

Yang–Mills measure and the master field on the sphere

Antoine Dahlqvist¹, James R. Norris²

¹ School of Mathematical and Physical Sciences, University of Sussex, Brighton, BN1 9RH, UK

E-mail: a.dahlqvist@sussex.ac.uk

² Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WB, UK

E-mail: j.r.norris@statslab.cam.ac.uk

Received: date / Accepted: date

Abstract: We study the Yang–Mills measure on the sphere with unitary structure group. In the limit where the structure group has high dimension, we show that the traces of loop holonomies converge in probability to a deterministic limit, which is known as the master field on the sphere. The values of the master field on simple loops are expressed in terms of the solution of a variational problem. We show that, given its values on simple loops, the master field is characterized on all loops of finite length by a system of differential equations, known as the Makeenko–Migdal equations. We obtain a number of further properties of the master field. On specializing to families of simple loops, our results identify the high-dimensional limit, in non-commutative distribution, of the Brownian bridge in the group of unitary matrices starting and ending at the identity.

Contents

1. Introduction	2
2. Setting and statement of the main results	3
2.1 Yang–Mills measure on a compact Riemann surface	4
2.2 Embedded Brownian loops	6
2.3 Convergence to the master field on the sphere	6
2.4 Characterization of the master field on the sphere	7
2.5 Outline of the main argument	10
2.6 Convergence of spectral measures	11
2.7 Free unitary Brownian loop	12
2.8 The master field as a holonomy in $U(\infty)$	13
3. Harmonic analysis in $U(N)$ and a discrete β -ensemble	17
3.1 A representation formula	17
3.2 Concentration for the discrete β -ensemble and tightness of the support	21
3.3 Dimension-free continuity estimate for the holonomy of a simple loop	23
3.4 Evaluation of some contour integrals	24
3.5 Proof of Proposition 2.5	26
4. Makeenko–Migdal equations	28
4.1 Combinatorial planar graphs and loops	28
4.2 Generalized Makeenko–Migdal equations	30

4.3	Makeenko–Migdal equations for Wilson loops	31
4.4	Makeenko–Migdal vectors and the winding number	33
4.5	Proof of Proposition 2.6	35
5.	Extension to loops of finite length	39
5.1	Some estimates for piecewise geodesic loops	39
5.2	Proof of Proposition 2.7	42
6.	Further properties of the master field	44
6.1	Relation with the Hermitian Brownian loop	44
6.2	Duality at the midpoint of the loop	45
6.3	Convergence to the planar master field	45
6.4	Uniqueness of the master field	47
6.5	Combinatorial formulas for the master field	50

1. Introduction

The Yang–Mills measure, associated to a (two-dimensional) surface Σ and to a compact Lie group G , is a probability measure on (generalized) connections of principal G -bundles over Σ . It was introduced in a series of works by Gross, King and Sengupta [29], Fine [21], Driver [17], Witten [54, 55], Sengupta [50] and Lévy [39], as a mathematical version of Euclidean Yang–Mills field theory. See also [11] for recent progress in higher dimensions. In this paper, we will consider the Yang–Mills measure in the case where the surface Σ is fixed and the group G is a classical matrix group of high dimension. The interest of such a set-up from the viewpoint of random matrix theory was first raised in the mathematics literature by Singer [52], who made several conjectures, based on earlier work in physics [26, 27, 35, 36]. The high-dimensional limit of the Yang–Mills measure when Σ is the whole plane has since been studied by Xu [56], Sengupta [51], Lévy [41], Anshelevich and Sengupta [1], Dahlqvist [13] and others [8, 24].

We focus here on the case where the surface Σ is a sphere. This case has received particular attention in the physics literature [6, 14, 28, 49] as it displays a phase transition of third order named after Douglas and Kazakov [15]. A corresponding mathematical analysis of the partition function was achieved by Boutet de Monvel and Shcherbina [7] and Lévy and Maïda [44]. The main result of the present work, Theorem 2.2, confirms a conjecture of Singer [52], showing that, under the Yang–Mills measure on the sphere for the unitary group $U(N)$, the traces of loop holonomies converge as $N \rightarrow \infty$ to a deterministic limit. We characterize this limit analytically and derive some further properties. Following the physics literature, the limit is called the master field on the sphere.

As a by-product of our main result, we show that the Brownian loop in $U(N)$, that is to say, the Brownian bridge starting and ending at the identity, converges in non-commutative distribution as $N \rightarrow \infty$ to a certain non-commutative process, which we call the free unitary Brownian loop. The notion of free unitary Brownian motion was first defined by Biane [4], using free stochastic calculus, and moreover identified as a limit of the Brownian motion on $U(N)$. This latter limit was further studied in Lévy [40], Lévy and Maïda [43] and Collins, Dahlqvist and Kemp [12]. Our work may be considered as a first instance of the free unitary Brownian loop as a limit of a matrix valued process. Defining it directly in the setting of free probability is an interesting open problem that cannot be handled so far by classical tools of free stochastic calculus, such as the ones introduced in [4, 5].

There is a system of relations, discovered by Makeenko and Migdal [46], indexed by families of embedded loops, between the expectations under the Yang–Mills measure of polynomials in the traces of loop holonomies. These have now been proved for the whole plane by Lévy [41] and Dahlqvist [13] and for any compact surface by Driver, Gabriel, Hall and Kemp [18]. They belong to the class of Schwinger–Dyson equations, a family of equations obtained by generalizing integration-by-parts formulas to the setting of functional integrals. See for example [30] and [9, 10], where these equations are proved and used in different models of random unitary matrices and for

a lattice version of the Yang–Mills measure. For the Yang–Mills functional integral, this heuristic derivation has been justified recently by Driver [16]. The Makeenko–Migdal equations provide a potential line of argument to prove convergence of the Yang–Mills measure as $N \rightarrow \infty$, which is to show a suitable concentration estimate for the holonomy traces, and to pass to the limit in the equations, showing that the limit equations determine a unique limit object. In the whole plane case, moment estimates for unitary Brownian motion provide the needed concentration, and the Makeenko–Migdal equations may be augmented by a further equation, such that the whole system of equations then characterizes the limit field. So the programme has been completed in that case [13, 41]. However, as noted in [18], the concentration and characterization problems have remained open in general.

In this paper, we will establish two key points. First, for simple loops, we show in Proposition 3.1 that expectations and covariances of the holonomy traces can be represented by functionals of a discrete β -ensemble. This representation allows us to identify the limit in probability of these traces as $N \rightarrow \infty$, following the work of Guionnet and Maïda [30], Johansson [32] and Féral [20] on discrete β -ensembles. This amounts to a rigorous version of ideas explained by Boulatov [6] and Douglas and Kazakov [15]. The second point, shown in Section 4 using the Makeenko–Migdal equations, is that the convergence of marginals to a deterministic limit for simple loops forces the same to hold for a more general class of loops¹. Then, by adapting some estimates of Lévy [41], we are able to consider eventually all loops of finite length, allowing us to express certain key properties of the master field in a natural way.

An alternative line of argument for the first point, which we shall discuss elsewhere, would be to use the fact that the process of eigenvalues of the marginals of the Brownian loop is known to have the same law as a Dyson Brownian motion on the circle, starting from 1 and conditioned to return to 1. Indeed, several scaling limits of this conditioned process have recently been understood by Liechty and Wang [45]. This link was first observed in the physics literature by Forrester, Majumdar and Schehr [22, 23]. Section 3 gives another way to obtain macroscopic results on the empirical distribution of this process.

The paper is organized as follows. Section 2 introduces the model and our results. Section 3 shows convergence and concentration of holonomy traces for simple loops, using a duality relation with a discrete β -ensemble. Section 4 explains how the Makeenko–Migdal equations can be used to extend this convergence to a general class of regular loops. Then, in Section 5, we make a final extension to all loops of finite length. Section 6 presents some further properties of the master field, including a relation with the free Hermitian Brownian loop in the subcritical regime, and a formula for the evaluation of the master field on a large class of loops.

Subject to certain modifications, to be explained in a future work, the argument explained here applies to other series of compact groups and also with the projective plane in place of the sphere.

2. Setting and statement of the main results

We review the notion of a Yang–Mills holonomy field over a compact Riemann surface. Then we discuss its relation, in the case of the sphere, to the Brownian loop in a Lie group. Next, we state our main results on convergence of Yang–Mills holonomy in $U(N)$ over the sphere to the master field, and on analytic characterization of the master field. The proof of these main results has three steps, which are outlined in Subsection 2.5. Then we discuss some consequences of our results, for the convergence of spectral measures of loop holonomies, and for the high-dimensional limit of the Brownian loop in $U(N)$. Finally, we discuss how the master field can be considered as a natural family of infinite-dimensional unitary transport operators, following up some suggestions of Singer [52].

¹ This point has recently been shown independently also by Brian Hall [31].

2.1. Yang–Mills measure on a compact Riemann surface. We recall in this subsection the notion of Yang–Mills measure in two dimensions, following the formulation of Lévy [39], as a field of holonomies indexed by paths of finite length. Let Σ be a two-dimensional Riemannian manifold and let G be a compact Lie group. Fix an area measure on Σ having a continuous positive density with respect to Lebesgue measure in each coordinate chart. Write T for the total area of Σ and denote by 1 the unit element of G . Fix a bi-invariant Riemannian metric on G and denote the associated heat kernel by $p = (p_t(g) : t \in (0, \infty), g \in G)$. Thus p is the unique C^∞ positive function on $(0, \infty) \times G$ such that

$$\frac{\partial p}{\partial t} = \frac{1}{2} \Delta p$$

and, for all continuous functions f on G , in the limit $t \rightarrow 0$,

$$\int_G f(g) p_t(g) dg \rightarrow f(1).$$

Here we have written Δ for the Laplace–Beltrami operator and dg for the normalized Haar measure on G .

We specialize in later sections to the case where Σ is the Euclidean sphere \mathbb{S}_T of total area T

$$\mathbb{S}_T = \{x \in \mathbb{R}^3 : 4\pi|x|^2 = T\}$$

and where G is the group $U(N)$ of unitary $N \times N$ matrices. The Lie algebra of $U(N)$ is the space of skew-Hermitian matrices $\mathfrak{u}(N)$. We specify a metric on $U(N)$ by the following choice of inner product on $\mathfrak{u}(N)$

$$\langle g_1, g_2 \rangle = N \operatorname{Tr}(g_1 g_2^*) \quad (1)$$

where $\operatorname{Tr}(g) = \sum_{i=1}^N g_{ii}$. This dependence of the metric on N , which is standard in random matrix theory, is chosen so that the objects of interest to us have a non-trivial scaling limit as $N \rightarrow \infty$.

By an oriented path in Σ we mean a continuous map $[0, 1] \rightarrow \Sigma$. Write $\operatorname{Path}(\Sigma)$ for the set of oriented paths of finite length in Σ , parametrized by $[0, 1]$ at constant speed. Denote the length of a path $\gamma \in \operatorname{Path}(\Sigma)$ by $\ell(\gamma)$. We consider $\operatorname{Path}(\Sigma)$ as a metric space, with the *length metric*

$$d(\gamma, \gamma') = |\ell(\gamma) - \ell(\gamma')| + \inf_{\tau} \sup_{t \in [0, 1]} d(\gamma_{\tau(t)}, \gamma'_t) \quad (2)$$

where the infimum is taken over homeomorphisms τ of $[0, 1]$. Each path γ has a starting point $\underline{\gamma}$ and a terminal point $\bar{\gamma}$. Write γ^{-1} for the reversal of γ , that is, the path of reverse orientation from $\bar{\gamma}$ to $\underline{\gamma}$. For paths γ_1, γ_2 such that $\bar{\gamma}_1 = \underline{\gamma}_2$, we write $\gamma_1 \gamma_2$ for the path obtained by their concatenation (and reparametrization by $[0, 1]$ at constant speed). Write $\operatorname{Loop}(\Sigma)$ for the set of loops of finite length in Σ . Thus

$$\operatorname{Loop}(\Sigma) = \{\gamma \in \operatorname{Path}(\Sigma) : \underline{\gamma} = \bar{\gamma}\}.$$

Write also $\operatorname{Path}_{x,y}(\mathbb{S}_T)$ for the set of paths from x to y , and $\operatorname{Loop}_x(\mathbb{S}_T)$ for the set of loops based at x . Given paths γ, γ_0 , we say that γ_0 is a *simple reduction* of γ if we can write γ and γ_0 as concatenations

$$\gamma = \gamma_1 \gamma_* \gamma_*^{-1} \gamma_2, \quad \gamma_0 = \gamma_1 \gamma_2$$

for some paths $\gamma_1, \gamma_2, \gamma_*$. More generally, we say that γ_0 is a *reduction* of γ if there is a sequence of paths $(\gamma_1, \dots, \gamma_n)$ such that γ_{i-1} is a simple reduction of γ_i for all i and $\gamma_n = \gamma$. Given paths γ_1, γ_2 , we write $\gamma_1 \sim \gamma_2$ if there is a path γ_0 which is a reduction of both γ_1 and γ_2 .

Given a subset Γ of $\operatorname{Path}(\Sigma)$ which is closed under reversal and concatenation, we call a function $h : \Gamma \rightarrow G$ *multiplicative* if

$$h_{\gamma^{-1}} = h_{\gamma}^{-1}, \quad h_{\gamma_1 \gamma_2} = h_{\gamma_2} h_{\gamma_1}$$

for all γ and for all γ_1, γ_2 with $\bar{\gamma}_1 = \gamma_2$. We denote the set of such multiplicative functions by $\text{Mult}(\Gamma, G)$. Note that, for any such function h , we have $h_{\gamma_1} = h_{\gamma_2}$ whenever $\gamma_1 \sim \gamma_2$.

A path is simple if it is injective on $[0, 1]$, while a loop is simple if it is injective as a map on the circle. We say that a finite subset $\mathbb{G} = \{e_1, \dots, e_m\} \subseteq \text{Path}(\Sigma)$ is an *embedded graph* in Σ if each path e_j is non-constant, is either simple or a simple loop, and meets other paths e_k only at its endpoints. Then we refer to the sequence (e_1, \dots, e_m) as a *labelled embedded graph*. We will sometimes write abusively $\mathbb{G} = (V, E, F)$ to mean that V is the set of endpoints of paths in \mathbb{G} , $E = \mathbb{G}$ and F is the set of connected components of $\Sigma \setminus \{e^* : e \in \mathbb{G}\}$. Here

$$e^* = \{e(t) : t \in [0, 1]\}.$$

We say that an embedded graph \mathbb{G} is a *discretization* of Σ if each face $f \in F$ is a simply connected domain in Σ . Write $\text{Path}(\mathbb{G})$ for the subset of $\text{Path}(\Sigma)$ obtained by concatenations of the paths in \mathbb{G} and their reversals.

A random process $H = (H_\gamma : \gamma \in \text{Path}(\Sigma))$ (on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$) taking values in G is a *Yang–Mills holonomy field* if

- (a) H is multiplicative, that is, $H(\omega) \in \text{Mult}(\text{Path}(\Sigma), G)$ for all $\omega \in \Omega$,
- (b) for any discretization $\mathbb{G} = (V, E, F)$ of Σ and all $h \in \text{Mult}(\text{Path}(\mathbb{G}), G)$,

$$\mathbb{P}(H_e \in dh_e \text{ for all } e \in E) = p_T(1)^{-1} \prod_{f \in F} p_{\text{area}(f)}(h_f) \prod_{e \in E} dh_e \quad (3)$$

- (c) for any convergent sequence $\gamma(n) \rightarrow \gamma$ in $\text{Path}(\Sigma)$ with fixed endpoints,

$$H_{\gamma(n)} \rightarrow H_\gamma \quad \text{in probability.} \quad (4)$$

The equation (3) specifies certain finite-dimensional distributions of H , considered as probability measures on G^E . The volume element $\prod_{e \in E} dh_e$ is the product of normalized Haar measures on G . For each face f , we have chosen a simple loop $\gamma(f) \in \text{Loop}(\mathbb{G})$ whose range is the boundary of f and set $h_f = h_{\gamma(f)}$. The invariance properties of Haar measure and the heat kernel under inversion and conjugation guarantee that the expression (3) depends neither on the orientations of the edges nor on the choice of loops bounding the faces.

For each $\gamma \in \text{Path}(\Sigma)$, we can define a coordinate function $H_\gamma : \text{Mult}(\text{Path}(\Sigma), G) \rightarrow G$ by $H_\gamma(h) = h_\gamma$. We define a σ -algebra \mathcal{C} on $\text{Mult}(\text{Path}(\Sigma), G)$ by

$$\mathcal{C} = \sigma(H_\gamma : \gamma \in \text{Path}(\Sigma)).$$

Then $(H_\gamma : \gamma \in \text{Path}(\Sigma))$ is a multiplicative random process on $(\text{Mult}(\text{Path}(\Sigma), G), \mathcal{C})$. We use the same notation $(H_\gamma : \gamma \in \text{Path}(\Sigma))$ both for this canonical coordinate process and also, more generally, for any multiplicative random process.

Our basic object of study is the *Yang–Mills measure* provided by the following theorem of Lévy [39, Theorem 2.62], building on earlier work of Driver [17] and Sengupta [50].

Theorem 2.1 *There is a unique probability measure on $(\text{Mult}(\text{Path}(\Sigma), G), \mathcal{C})$ under which the coordinate process $(H_\gamma : \gamma \in \text{Path}(\Sigma))$ is a Yang–Mills holonomy field.*

Let $H = (H_\gamma : \gamma \in \text{Path}(\Sigma))$ be a Yang–Mills holonomy field in G . We note the following properties of *gauge invariance* and *invariance under area-preserving diffeomorphisms*, which follow from invariance properties of (3) and the uniqueness statement of the theorem. Let $s : \Sigma \rightarrow G$ be any function and let $\psi : \Sigma \rightarrow \Sigma$ be an area-preserving diffeomorphism. Consider the processes

$$H^s = (s(\bar{\gamma})H_\gamma s(\gamma)^{-1} : \gamma \in \text{Path}(\Sigma)), \quad H^\psi = (H_{\psi \circ \gamma} : \gamma \in \text{Path}(\Sigma)).$$

Then H^s and H^ψ have the same law as H . In particular, the relevant data from Σ are just its genus and the total area T .

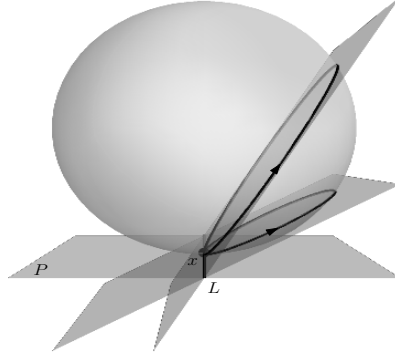


Fig. 1. A family of simple loops obtained by intersecting the sphere with a rotating plane, as described in Subsection 2.2.

2.2. Embedded Brownian loops. We specialize now to the case where the surface Σ is the sphere \mathbb{S}_T of area T . In each Yang–Mills holonomy field $H = (H_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$, there are many embedded Brownian loops in G based at 1 and parametrized by $[0, T]$, as we now show. Recall that a random process $B = (B_t : t \in [0, T])$ taking values in G is a Brownian loop based at 1 if

- (a) B is continuous, that is, $B(\omega) \in C([0, T], G)$ for all $\omega \in \Omega$,
- (b) for all $n \in \mathbb{N}$, all $g_1, \dots, g_{n-1} \in G$ and all increasing sequences (t_1, \dots, t_{n-1}) in $(0, T)$, setting $g_0 = g_n = 1$ and $t_0 = 0$ and $t_n = T$ and writing $t_k = s_1 + \dots + s_k$,

$$\mathbb{P}(B_{t_k} \in dg_k \text{ for } k = 1, \dots, n-1) = p_T(1)^{-1} \prod_{i=1}^n p_{s_i}(g_i g_{i-1}^{-1}) \prod_{k=1}^{n-1} dg_k.$$

Choose a point x in \mathbb{S}_T and let P be a tangent plane to \mathbb{S}_T at x , considered as embedded in \mathbb{R}^3 . Choose a line L in P through x and rotate P once around L . The resulting intersections of P with \mathbb{S}_T , which are a nested family of circles, may be given a consistent orientation and then considered as a family in $\text{Loop}(\mathbb{S}_T)$, all based at x . We can parametrize this family of loops as $(l(t) : t \in [0, T])$ so that the domain inside $l(t)$ has area t for all T . Then, for all $n \in \mathbb{N}$ and all sequences (t_1, \dots, t_{n-1}) in $(0, T)$, the loops $l(t_1), \dots, l(t_{n-1})$ are the edges of a discretization of \mathbb{S}_T . Define a random process $\beta = (\beta_t : t \in [0, T])$ in G by

$$\beta_t = H_{l(t)}.$$

It is straightforward to deduce from property (b) of the Yang–Mills holonomy field that the finite-dimensional distributions of β satisfy condition (b) for the Brownian loop. Hence, by standard arguments, β has a continuous version, B say, which is a Brownian loop in G based at 1. The reader will see many ways to vary this construction while still obtaining a Brownian loop.

2.3. Convergence to the master field on the sphere. We specialize now to the case where the structure group G is the group of $N \times N$ unitary matrices $U(N)$. Let $H^N = (H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ be a Yang–Mills holonomy field in $U(N)$ over the sphere \mathbb{S}_T of area T . Our main results establish a law of large numbers for this random field in the limit $N \rightarrow \infty$, which we express for now in terms of the normalized trace

$$\text{tr}_N(g) = \text{tr}(g) = N^{-1} \sum_{i=1}^N g_{ii}.$$

Here is our first main result.

Theorem 2.2 *There exists a function on loops*

$$\Phi_T : \text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$$

such that, for all $l \in \text{Loop}(\mathbb{S}_T)$,

$$\text{tr}_N(H_l^N) \rightarrow \Phi_T(l) \quad \text{in probability as } N \rightarrow \infty.$$

The function Φ_T is known in the physics literature as the *master field on the sphere*. Until we have proved Theorem 2.2, it will be convenient provisionally to define Φ_T by

$$\Phi_T(l) = \begin{cases} \lim_{N \rightarrow \infty} \mathbb{E}(\text{tr}_N(H_l^N)), & \text{if this limit exists,} \\ 0, & \text{otherwise.} \end{cases}$$

Note that, since $|\text{tr}_N(H_l^N)| \leq 1$, by bounded convergence, as soon as we show that $\text{tr}_N(H_l^N)$ converges in probability with deterministic limit, it will follow that $\mathbb{E}(\text{tr}_N(H_l^N))$ converges with the same limit, so that limit must equal $\Phi_T(l)$ as provisionally defined.

Given Theorem 2.2, the master field inherits certain properties from its finite- N approximations $\mathbb{E}(\text{tr}_N(H_l^N))$, as the reader may easily check.

Proposition 2.3 *The master field Φ_T has the following properties:*

- (a) $\Phi_T = 1$ on constant loops and $\Phi_T(l) = \Phi_T(l^{-1}) \in [-1, 1]$ for all loops l ,
- (b) $\Phi_T(\gamma_1 \gamma_2) = \Phi_T(\gamma_2 \gamma_1)$ for all pairs of paths γ_1, γ_2 such that $\gamma_1 \gamma_2$ is a loop,
- (c) $\Phi_T(l_1) = \Phi_T(l_2)$ whenever $l_1 \sim l_2$,
- (d) for all $x, y \in \mathbb{S}_T$, all $n \in \mathbb{N}$, all $a_1, \dots, a_n \in \mathbb{C}$ and all $\gamma_1, \dots, \gamma_n \in \text{Path}_{x,y}(\mathbb{S}_T)$,

$$\sum_{i,j=1}^n a_i \overline{a_j} \Phi_T(\gamma_i \gamma_j^{-1}) \geq 0$$

- (e) for all loops l and any area-preserving diffeomorphism ψ of \mathbb{S}_T ,

$$\Phi_T(\psi(l)) = \Phi_T(l).$$

2.4. Characterization of the master field on the sphere. Our second main result is an analytic characterization of the master field. This will require some associated notions which we now introduce. Consider the following variational problem: minimize the functional

$$\mathcal{I}_T(\mu) = \int_{\mathbb{R}^2} \left\{ \frac{1}{2}(x^2 + y^2)T - 2 \log |x - y| \right\} \mu(dx) \mu(dy) \quad (5)$$

over the set of probability measures μ on \mathbb{R} such that, for all intervals $[a, b]$,

$$\mu([a, b]) \leq b - a.$$

We note for later use some statements concerning this problem, proofs of which may be found in Lévy and Maïda [44]. First, the functional \mathcal{I}_T is well-defined on the given set of probability measures, with values in $(-\infty, \infty]$, and has a unique minimizer, which we denote by μ_T . Then μ_T has a continuous density function ρ_T with respect to Lebesgue measure, with $0 \leq \rho_T(x) \leq 1$ for all x . In the case $T \in (0, \pi^2]$, ρ_T is the semi-circle density of variance $1/T$, given by

$$\rho_T(x) = \frac{T}{2\pi} \sqrt{\frac{4}{T} - x^2}, \quad |x| \leq \frac{2}{\sqrt{T}}. \quad (6)$$

Note that the right-hand side in (6) exceeds 1 when $x = 0$ for $T > \pi^2$.

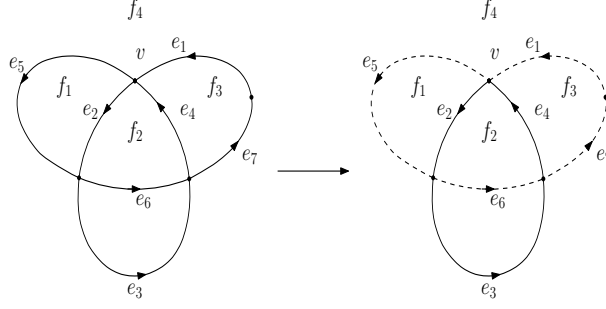


Fig. 2. The splitting of the loop $e_1e_2\dots e_7$ on the left-hand-side yields the two loops $e_2e_3e_4$ and $e_5e_6e_7e_1$ respectively represented with plain and dashed strands.

For $T \in (\pi^2, \infty)$, there is a unique $k \in (0, 1)$ such that

$$T = 8EK - 4(1 - k^2)K^2$$

where $K = K(k)$ and $E = E(k)$ are, respectively, the complete elliptic integrals of the first and second kind. See for example [38, Chapter 3, equations (3.1.3) and (3.5.4)]. Set

$$\alpha = 4kK/T, \quad \beta = 4K/T. \quad (7)$$

Then the minimizing density ρ_T is identically 1 on $[-\alpha, \alpha]$, is supported on $[-\beta, \beta]$, and satisfies, for $|x| \in (\alpha, \beta)$,

$$\rho_T(x) = \frac{2\sqrt{(x^2 - \alpha^2)(\beta^2 - x^2)}}{\pi\beta|x|} \int_0^1 \frac{ds}{(1 - \alpha^2 s^2/x^2)\sqrt{(1 - s^2)(1 - \alpha^2 s^2/\beta^2)}}. \quad (8)$$

See [44, Lemma 4.7, equation (4.14)]. See also [44, Figure 7] for an informative plot of the family of densities $(\rho_T : T \in (0, \infty))$.

Let us say that $l \in \text{Loop}(\mathbb{S}_T)$ is a *regular loop* if there is a labelled embedded graph $\mathbb{G}_l = (e_1, \dots, e_m)$ in $\text{Path}(\mathbb{S}_T)$ such that l is given by the concatenation $e_1 \dots e_m$, in which e_1 has degree 2 and in which e_2, \dots, e_m have degree 4 and are transverse self-intersections of l . Here, we say that a self-intersection of l at a vertex v of degree 4 is *transverse* if, as l passes through v , it arrives and leaves by opposite edges. Note that \mathbb{G}_l is then uniquely determined by l .

Given a regular loop l and a point v of self-intersection of l , there are two regular loops l_v and \hat{l}_v starting from v , obtained by *splitting* l at v , that is, by following l on its first and second exit from v , respectively, until it first returns to v . Note that both l_v and \hat{l}_v have fewer self-intersections than l . See Figure 2 for an example. Denote by f_1 the face of \mathbb{G} which is adjacent to the two outgoing strands of l at v , and denote by f_1, f_2, f_3, f_4 the faces of \mathbb{G} found on making a small circuit around v in the positive sense, starting from f_1 . Note that these faces may not all be distinct. For each face f of \mathbb{G} , define

$$\text{sgn}_v(f) = \sum_{k=1}^4 (-1)^{k+1} 1_{\{f_k\}}(f).$$

In the case where the faces f_1, f_2, f_3, f_4 are distinct, we have $\text{sgn}_v(f_1) = \text{sgn}_v(f_3) = 1$ and $\text{sgn}_v(f_2) = \text{sgn}_v(f_4) = -1$. Since \mathbb{G} is embedded in the sphere, the only other possibility is that $f_1 = f_3 \neq f_2 \neq f_4 \neq f_1$, in which case $\text{sgn}_v(f_1) = 2$ and $\text{sgn}_v(f_2) = \text{sgn}_v(f_4) = -1$. See Figure 3. For $\eta > 0$, we say that a C^∞ map

$$\theta : [0, \eta) \times \mathbb{S}_T \rightarrow \mathbb{S}_T$$

is a *Makeenko–Migdal flow* at (l, v) if

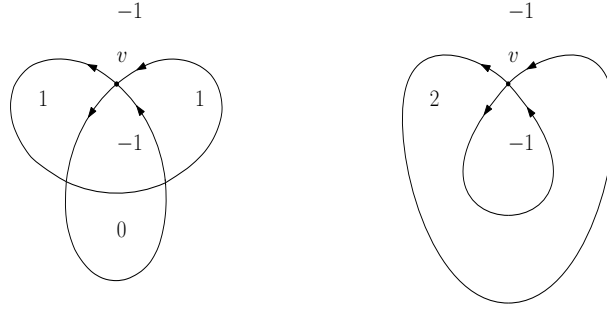


Fig. 3. The value of sgn_v , where v is denoted by a dot, is printed on each face of the embedded graph for two different loops.

- (a) $\theta(0, x) = x$ for all x ,
- (b) $\theta(t, \cdot)$ is a diffeomorphism of \mathbb{S}_T for all t ,
- (c) for any face f of the embedded graph \mathbb{G} ,

$$\frac{d}{dt} \text{area}(\theta(t, f)) = \text{sgn}_v(f). \quad (9)$$

We can now state our analytic characterization of the master field.

Theorem 2.4 *The master field $\Phi_T : \text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$ has the following properties, which together characterize it uniquely:*

- (a) Φ_T is continuous with respect to the length metric on $\text{Loop}(\mathbb{S}_T)$,
- (b) Φ_T is invariant under reduction: for all pairs of loops l_1, l_2 with $l_1 \sim l_2$,

$$\Phi_T(l_1) = \Phi_T(l_2)$$

- (c) Φ_T is invariant under area-preserving homeomorphisms: for all regular loops l and any area-preserving homeomorphism ξ of \mathbb{S}_T such that $\xi(l) \in \text{Loop}(\mathbb{S}_T)$,

$$\Phi_T(\xi(l)) = \Phi_T(l)$$

- (d) Φ_T satisfies the Makeenko–Migdal equations: for all regular loops l , all points v of self-intersection of l , and any Makeenko–Migdal flow θ at (l, v) , $\Phi_T(\theta(t, l))$ is differentiable in t at 0 with

$$\left. \frac{d}{dt} \right|_{t=0} \Phi_T(\theta(t, l)) = \Phi_T(l_v) \Phi_T(\hat{l}_v) \quad (10)$$

- (e) for all simple loops l and all $n \in \mathbb{N}$,

$$\Phi_T(l^n) = \frac{2}{n\pi} \int_0^\infty \cosh\{(a_1 - a_2)nx/2\} \sin\{n\pi\rho_T(x)\} dx \quad (11)$$

where a_1 and a_2 are the areas of the connected components of $\mathbb{S}_T \setminus l^*$.

Note that the integrand in (11) vanishes whenever $\rho_T(x) = 0$ or $\rho_T(x) = 1$. In property (e), we have written l^n for the n -fold concatenation of l with itself. In fact, it suffices for uniqueness that property (e) hold in the case $n = 1$, as we show in Subsection 6.4.

2.5. Outline of the main argument. We now outline the main steps in our proof of Theorems 2.2 and 2.4. We build progressively an understanding of the limit, first for simple loops, then regular loops, and finally for all loops of finite length. First, we prove in Subsection 3.5 the following statement for simple loops. The argument uses harmonic analysis in $U(N)$ to express means and covariances of $\text{tr}_N(H_{l^n}^N)$ in terms of a discrete β -ensemble, whose asymptotics as $N \rightarrow \infty$ we can compute. Write $\text{Loop}_0(\mathbb{S}_T)$ for the set of simple loops in $\text{Loop}(\mathbb{S}_T)$. Let $l \in \text{Loop}_0(\mathbb{S}_T)$ and recall that we write l^* for the range of l . Then $\mathbb{S}_T \setminus l^*$ has two connected components. Write $a_1(l)$ for the area of the component on the left of l and $a_2(l)$ for the area of the component on the right. Then $a_1(l), a_2(l) > 0$ and $a_1(l) + a_2(l) = T$. Set

$$\phi_T(n, a_1, a_2) = \frac{2}{n\pi} \int_0^\infty \cosh\{(a_1 - a_2)nx/2\} \sin\{n\pi\rho_T(x)\} dx. \quad (12)$$

Proposition 2.5 *For all $n \in \mathbb{N}$,*

$$\text{tr}_N(H_{l^n}^N) \rightarrow \Phi_T(l^n) = \phi_T(n, a_1(l), a_2(l))$$

uniformly in $l \in \text{Loop}_0(\mathbb{S}_T)$ in $L^2(\mathbb{P})$ as $N \rightarrow \infty$.

The next step is the following proposition, which is proved in Subsection 4.5. The argument is based on the Makeenko–Migdal equations for Wilson loops, which will be discussed in Subsection 4.3. Write $\text{Loop}_n(\mathbb{S}_T)$ the set of regular loops having at most n self-intersections.

In order to state the proposition, we will need to introduce certain quantities associated to a regular loop $l \in \text{Loop}_n(\mathbb{S}_T)$. There is a winding number function $n_l : \mathbb{S}_T \setminus l^* \rightarrow \mathbb{Z}$, which we fix uniquely by requiring its minimal value to be 0. The winding number is discussed in more detail in Subsection 4.4. The function n_l is constant on the faces of the associated embedded graph, which are the connected components of $\mathbb{S}_T \setminus l^*$. The notion of continuity in area is defined at the end of section 4.1.

Proposition 2.6 *For all $n \in \mathbb{N}$,*

$$\text{tr}_N(H_l^N) \rightarrow \Phi_T(l)$$

uniformly in $l \in \text{Loop}_n(\mathbb{S}_T)$ in $L^2(\mathbb{P})$ as $N \rightarrow \infty$. Moreover, the restriction of the master field Φ_T to $\text{Loop}_n(\mathbb{S}_T)$ is the unique continuous function $\text{Loop}_n(\mathbb{S}_T) \rightarrow \mathbb{C}$ with the following properties: it is invariant under area-preserving homeomorphisms and uniformly continuous in area, it satisfies the Makeenko–Migdal equations (10), and satisfies, for some constant $C_n < \infty$ and all loops $l \in \text{Loop}_n(\mathbb{S}_T)$,

$$|\Phi_T(l) - \phi_T(n_*, a_0, a_*)| \leq C_n(T - a_{k_0} - a_{k_*}) \quad (13)$$

where n_ is the maximum of the winding number function n_l , where k_0 and k_* are the indices of faces of minimal and maximal winding number, and where a_0 and a_* are determined by*

$$a_0 + a_* = T, \quad a_* n_* = \sum_{k=1}^p a_k n_k$$

where a_k and n_k , for $k = 1, \dots, p$, are respectively the area and the winding number of the face of index k .

Finally, we extend to all loops of finite length in the following proposition, which combines the statements of Theorems 2.2 and 2.4. The proof is given in Section 5, using approximation by piecewise geodesics, and by adapting some general arguments of Lévy [41].

Proposition 2.7 *For all $l \in \text{Loop}(\mathbb{S}_T)$,*

$$\text{tr}_N(H_l^N) \rightarrow \Phi_T(l)$$

in probability as $N \rightarrow \infty$. Moreover, the master field Φ_T is the unique continuous function $\text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$ with the following properties: it is invariant under reduction, invariant under area-preserving homeomorphisms, satisfies the Makeenko–Migdal equations (10) on regular loops, and satisfies (11) for simple loops.

2.6. Convergence of spectral measures. Let $(H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ be a Yang–Mills holonomy field in $U(N)$. For $l \in \text{Loop}(\mathbb{S}_T)$, consider the empirical eigenvalue distribution on the unit circle \mathbb{U} , given by

$$\nu_T^N(l) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$$

where $\lambda_1, \dots, \lambda_N$ are the eigenvalues of H_l^N enumerated with multiplicity. Write $\mathcal{M}_1(\mathbb{U})$ for the set of Borel probability measures on \mathbb{U} .

Corollary 2.8 *There is a function $\nu_T : \text{Loop}(\mathbb{S}_T) \rightarrow \mathcal{M}_1(\mathbb{U})$ such that, for all $l \in \text{Loop}(\mathbb{S}_T)$,*

$$\nu_T^N(l) \rightarrow \nu_T(l)$$

weakly in probability on \mathbb{U} as $N \rightarrow \infty$. Moreover, for all simple loops l and all $n \in \mathbb{N}$,

$$\int_{\mathbb{U}} \omega^n \nu_T(l)(d\omega) = \frac{2}{n\pi} \int_0^\infty \cosh\{(a_1(l) - a_2(l))nx/2\} \sin\{n\pi\rho_T(x)\} dx.$$

Moreover, for $T \in (0, \pi^2]$, all simple loops l , and all bounded Borel functions f ,

$$\int_{\mathbb{U}} f(\omega) \nu_T(l)(d\omega) = \int_{-\pi}^\pi f(e^{i\theta}) s_{a_1 a_2 / T}(\theta) d\theta \quad (14)$$

where s_t is the semi-circle density of variance t , given by

$$s_t(x) = \frac{1}{2\pi t} \sqrt{4t - x^2}, \quad |x| \leq 2\sqrt{t}. \quad (15)$$

Proof. By Theorem 2.2, for $l \in \text{Loop}(\mathbb{S}_T)$ and all $n \in \mathbb{N}$, we have

$$\int_{\mathbb{U}} \omega^n \nu_T^N(l)(d\omega) = \text{tr}(H_{l^n}^N) \rightarrow \Phi_T(l^n)$$

in probability as $N \rightarrow \infty$. Since \mathbb{U} is compact, by a standard tightness argument, it follows that there exists a probability measure $\nu_T(l)$ on \mathbb{U} such that

$$\int_{\mathbb{U}} \omega^n \nu_T(l)(d\omega) = \Phi_T(l^n)$$

for all $n \in \mathbb{N}$ and such that $\nu_T^N(l) \rightarrow \nu_T(l)$ weakly in probability as $N \rightarrow \infty$. By Theorem 2.4, $\Phi_T(l^n)$ is given by (11) for all simple loops l . Finally, we will show in Subsection 3.4 that, for all $T \in (0, \pi^2]$ and all $n \in \mathbb{N}$,

$$\Phi_T(l^n) = \int_{-\pi}^\pi e^{in\theta} s_{a_1 a_2 / T}(\theta) d\theta$$

so (14) holds for polynomials, and so it holds in general. \square

Thus, for $T \in (0, \pi^2]$ and for simple loops l , the limiting spectral measure $\nu_T(l)$ has a semi-circle density on \mathbb{U} , with

$$\text{supp}(\nu_T(l)) = \{e^{i\theta} : |\theta| \leq 2\sqrt{a_1 a_2 / T}\}.$$

The maximal support is then $\{e^{i\theta} : |\theta| \leq \sqrt{T}\}$, achieved when $a_1 = a_2 = T/2$. Note that, in the critical case $T = \pi^2$, the two endpoints of the maximal support meet at $\theta = \pm\pi$.

2.7. Free unitary Brownian loop. As a corollary of Theorem 2.2, we show that the Brownian loop in $U(N)$ based at 1 of lifetime T converges in non-commutative distribution as $N \rightarrow \infty$. Moreover, we identify the limiting empirical distribution of eigenvalues at each time $t \in [0, T]$.

Consider the free unital $*$ -algebra \mathcal{A}_T of polynomials over \mathbb{C} in the variables $(X_t : t \in [0, T])$ and their inverses. Each element $Q \in \mathcal{A}_T$ is a finite linear combination over \mathbb{C} of monomials of the form

$$X_{t_1}^{\varepsilon_1} \dots X_{t_n}^{\varepsilon_n}$$

where $t_1, \dots, t_n \in [0, T]$ and $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$. Thus each $Q \in \mathcal{A}_T$ may be written as a non-commutative polynomial

$$Q = q(X_t, X_t^{-1} : t \in [0, T])$$

with coefficients in \mathbb{C} . The operation $*$ is the unique conjugate-linear, anti-multiplicative involution on \mathcal{A}_T such that

$$X_t^* = X_t^{-1}.$$

For each $N \in \mathbb{N}$, there exists a Brownian loop $B^N = (B_t^N : t \in [0, T])$ in $U(N)$ based at 1 of parameter T . Define a random non-negative unit trace² on \mathcal{A}_T by setting

$$\tau_N(Q) = \text{tr}_N(q(B_t^N, (B_t^N)^{-1} : t \in [0, T])).$$

Theorem 2.9 *There is a non-negative unit trace τ_∞ on \mathcal{A}_T such that, for all $Q \in \mathcal{A}_T$,*

$$\tau_N(Q) \rightarrow \tau_\infty(Q) \quad \text{in probability as } N \rightarrow \infty.$$

Proof. It will suffice to consider the case where B^N is constructed from a Yang–Mills holonomy field $(H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ in $U(N)$, as in Subsection 2.2. Then, for some $x \in \mathbb{S}_T$ and some family of loops $l(t) \in \text{Loop}(\mathbb{S}_T)$ based at x , we have

$$B_t^N = H_{l(t)}^N \quad \text{almost surely, for all } t \in [0, T].$$

Consider first the case of a monomial $Q = X_{t_1}^{\varepsilon_1} \dots X_{t_n}^{\varepsilon_n}$ with $\varepsilon_1, \dots, \varepsilon_n \in \{-1, 1\}$ and set $l_Q = l(t_n)^{\varepsilon_n} \dots l(t_1)^{\varepsilon_1}$. Then, by Theorem 2.2,

$$\tau_N(Q) = \text{tr}((B_{t_1}^N)^{\varepsilon_1} \dots (B_{t_n}^N)^{\varepsilon_n}) = \text{tr}((H_{l(t_1)}^N)^{\varepsilon_1} \dots (H_{l(t_n)}^N)^{\varepsilon_n}) = \text{tr}(H_{l_Q}^N) \rightarrow \Phi_T(l_Q)$$

in probability as $N \rightarrow \infty$. Define $\tau_\infty(Q) = \Phi_T(l_Q)$ for all monomials Q and extend τ_∞ linearly to \mathcal{A}_T . Then $\tau_N(Q) \rightarrow \tau_\infty(Q)$ in probability as $N \rightarrow \infty$, for all $Q \in \mathcal{A}_T$, and τ_∞ inherits the property of being a non-negative unit trace from its random approximations τ_N . \square

Given a non-commutative random process $x = (x_t : t \in [0, T])$ in a non-commutative probability space (\mathcal{A}, τ) , let us say that x is a *free unitary Brownian loop* if, for all n , all $t_1, \dots, t_n \in [0, T]$ and all $(y_{t_k}, Y_{t_k}) \in \{(x_{t_k}, X_{t_k}), (x_{t_k}^*, X_{t_k}^*)\}$,

$$\tau(y_{t_1} \dots y_{t_n}) = \tau_\infty(Y_{t_1} \dots Y_{t_n}).$$

In particular, the canonical process $(X_t : t \in [0, T])$ is a free unitary Brownian loop in $(\mathcal{A}_T, \tau_\infty)$. We shall see in Section 6 that, in the subcritical regime $T \leq \pi^2$, a free unitary Brownian loop x has the same marginal distributions as e^{ib} , where b is a free Brownian loop with the same lifetime. Thus the spectral measure of each marginal of a free unitary Brownian loop is the push-forward of a Wigner law by the exponential mapping to the circle. However, we shall also see that the full non-commutative distributions of x and e^{ib} are different.

² Recall that a linear map τ on a unital $*$ -algebra \mathcal{A} is a *non-negative unit trace* if, for all $x, y \in \mathcal{A}$,

$$\tau(xx^*) \geq 0, \quad \tau(1) = 1, \quad \tau(xy) = \tau(yx).$$

The pair (\mathcal{A}, τ) is then a *non-commutative probability space*.

2.8. The master field as a holonomy in $U(\infty)$. We will carry out the suggestion of Singer [52], to use a variation of the Gelfand–Naimark–Segal construction to obtain from the master field a family of Hilbert spaces, indexed by \mathbb{S}_T , and equipped with a canonical connection, viewed as a family of unitary transport operators indexed by $\text{Path}(\mathbb{S}_T)$. First, in order to clarify and motivate this construction, we will make an analogous construction for finite N , showing its relation to the notion of Yang–Mills holonomy field in $U(N)$. Conditional on a certain non-degeneracy property for the master field, we will further exhibit the finite- N holonomy measures as recoverable by restriction of the limit holonomy field to certain invariant random subspaces.

We have presented the Yang–Mills holonomy field as a process $(H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ with values in $U(N)$. However, the property of gauge-invariance allows us to think of it as follows. Suppose we are given a family of complex vector spaces $V = (V_x : x \in \mathbb{S}_T)$, each equipped with a Hermitian inner product and having dimension N . Choose, for each $x \in \mathbb{S}_T$, a complex linear isometry $s(x) : \mathbb{C}^N \rightarrow V_x$. Given a Yang–Mills holonomy field $(H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ in $U(N)$, for each $\gamma \in \text{Path}_{x,y}(\mathbb{S}_T)$, we can define a complex linear isometry $T_\gamma : V_x \rightarrow V_y$ by

$$T_\gamma = s(y)H_\gamma^N s(x)^{-1}.$$

Then, by gauge invariance, the law of the process $(T_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ does not depend on the choice of the family of isometries $(s(x) : x \in \mathbb{S}_T)$. We call any process with this law a *Yang–Mills holonomy field in $\text{Isom}(V)$* . The original holonomy field $(H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ then corresponds to the case where $V_x = \mathbb{C}^N$ for all x . Moreover, given any Yang–Mills holonomy field $(T_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ in $\text{Isom}(V)$ and any choice of a family of complex linear isometries $s(x) : \mathbb{C}^N \rightarrow V_x$, we obtain a Yang–Mills holonomy field $(H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ in $U(N)$ by setting

$$H_\gamma^N = s(y)^{-1}T_\gamma s(x).$$

Proposition 2.10 *Let $(H_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ be a Yang–Mills holonomy field in $U(N)$ and let E be an independent uniformly random unit vector in \mathbb{C}^N . Define, for $l \in \text{Loop}(\mathbb{S}_T)$,*

$$\tau_T^N(l) = \langle E, H_l^N E \rangle.$$

Then $\tau_T^N(l) \rightarrow \Phi_T(l)$ in probability as $N \rightarrow \infty$ for all l .

We already know that

$$\mathbb{E}(\tau_T^N(l) | H^N) = \Phi_T^N(l) \rightarrow \Phi_T(l)$$

in probability as $N \rightarrow \infty$. The proposition thus shows that the same convergence in probability holds without taking the expectation over the random vector E . The extra randomness present in τ_T^N makes it a more natural object than Φ_T^N in certain constructions below.

Proof. Let Z_1, \dots, Z_N be independent complex Gaussian random variables. Set

$$S_N = \sum_{k=1}^N |Z_k|^2, \quad \tilde{E} = (Z_1, \dots, Z_N) / \sqrt{S_N}.$$

Then \tilde{E} is a uniform random unit vector in \mathbb{C}^N . Define, for $l \in \text{Loop}(\mathbb{S}_T)$,

$$\tilde{\tau}_T^N(l) = \sum_{k=1}^N \lambda_k |\tilde{E}_k|^2 = \sum_{k=1}^N \lambda_k |Z_k|^2 / S_N$$

where $\lambda_1, \dots, \lambda_N$ is an enumeration of the eigenvalues of H_l^N . Since H_l^N is diagonalized by unitary conjugation and \tilde{E} is independent of H_l^N , it follows that $\tilde{\tau}_T^N(l)$ has the same distribution as $\tau_T^N(l)$. By Corollary 2.8, the empirical distribution of eigenvalues

$$\nu_T^N(l) = \frac{1}{N} \sum_{k=1}^N \delta_{\lambda_k}$$

converges weakly in probability on the unit circle \mathbb{U} with deterministic limit $\nu_T(l)$ satisfying

$$\Phi_T(l) = \int_{\mathbb{U}} \lambda \nu_T(l)(d\lambda).$$

But $\mathbb{E}(|Z_k|^2) = 2$, so we obtain the following limits in L^2

$$\frac{1}{N} \sum_{k=1}^N \lambda_k |Z_k|^2 \rightarrow 2\Phi_T(l), \quad S_N/N \rightarrow 2.$$

Hence $\tilde{\tau}_T^N(l) \rightarrow \Phi_T(l)$ in probability as $N \rightarrow \infty$. \square

Fix a reference point $r \in \mathbb{S}_T$ and consider for each $x \in \mathbb{S}_T$ the vector space \mathcal{V}_x of complex functions on $\text{Path}_{r,x}(\mathbb{S}_T)$ of finite support. Thus, each $v \in \mathcal{V}_x$ has the form

$$v = \sum_{i=1}^n a_i \delta_{\gamma_i}$$

for some $n \geq 0$, with $a_i \in \mathbb{C}$ and $\gamma_i \in \text{Path}_{r,x}(\mathbb{S}_T)$ for all i . There are unique Hermitian forms $\langle \cdot, \cdot \rangle_x^N$ and $\langle \cdot, \cdot \rangle_x$ on \mathcal{V}_x such that, for all $\gamma_1, \gamma_2 \in \text{Path}_{r,x}(\mathbb{S}_T)$,

$$\langle \delta_{\gamma_1}, \delta_{\gamma_2} \rangle_x^N = \tau_T^N(\gamma_1 \gamma_2^{-1}), \quad \langle \delta_{\gamma_1}, \delta_{\gamma_2} \rangle_x = \Phi_T(\gamma_1 \gamma_2^{-1}).$$

By Proposition 2.10, $\langle v, v' \rangle_x^N \rightarrow \langle v, v' \rangle_x$ in probability for all v, v' . The form $\langle \cdot, \cdot \rangle_x^N$ is non-negative definite for all N , so $\langle \cdot, \cdot \rangle_x$ is also non-negative definite, as we observed in Proposition 2.3. For $x, y \in \mathbb{S}_T$ and $\gamma \in \text{Path}_{x,y}(\mathbb{S}_T)$, there is a unique complex linear map $T_\gamma : \mathcal{V}_x \rightarrow \mathcal{V}_y$ such that, for all $\gamma_0 \in \text{Path}_{r,x}(\mathbb{S}_T)$,

$$T_\gamma \delta_{\gamma_0} = \delta_{\gamma_0 \gamma}.$$

Note that, for $\gamma_1, \gamma_2 \in \text{Path}_{r,x}(\mathbb{S}_T)$,

$$\langle T_\gamma \delta_{\gamma_1}, T_\gamma \delta_{\gamma_2} \rangle_y^N = \langle \delta_{\gamma_1 \gamma}, \delta_{\gamma_2 \gamma} \rangle_y^N = \tau_T^N(\gamma_1 \gamma \gamma^{-1} \gamma_2^{-1}) = \tau_T^N(\gamma_1 \gamma_2^{-1}) = \langle \delta_{\gamma_1}, \delta_{\gamma_2} \rangle_x^N.$$

It follows that $\langle T_\gamma v_1, T_\gamma v_2 \rangle_y^N = \langle v_1, v_2 \rangle_x^N$ for all $v_1, v_2 \in \mathcal{V}_x$. Similarly, T_γ preserves the form $\langle \cdot, \cdot \rangle_x$.

For each $x \in \mathbb{S}_T$, write V_x^N for the quotient of the vector space \mathcal{V}_x by the kernel

$$\mathcal{K}_x^N = \{v \in \mathcal{V}_x : \langle v, v \rangle_x^N = 0\}.$$

Write $[v]^N = v + \mathcal{K}_x^N$. It is straightforward to check that, for $\gamma_1, \gamma_2 \in \text{Path}_{r,x}(\mathbb{S}_T)$ with $\gamma_1 \sim \gamma_2$, we have $[\delta_{\gamma_1}]^N = [\delta_{\gamma_2}]^N$ in V_x^N .

Proposition 2.11 *Almost surely, for all $x \in \mathbb{S}_T$, the random vector space V_x^N has finite dimension N .*

Proof. Since T_γ^N is a linear isometry for all paths $\gamma \in \text{Path}(\mathbb{S}_T)$, it will suffice to consider the case $x = r$. Let $(l(t) : t \in [0, T])$ be a family of loops in $\text{Loop}_x(\mathbb{S}_T)$ such as considered in Section 2.2, now with $x = r$. For $k = 1, \dots, N$ set $l_k = l((k-1)T/N)$. Write Ω_0 for the event that the random vectors $H_{l_1}E = E, H_{l_2}E, \dots, H_{l_N}E$ are linearly independent in \mathbb{C}^N . The joint law of H_{l_2}, \dots, H_{l_N} has a density ρ with respect to the product of normalized Haar measures on $U(N)^{N-1}$ given by

$$\rho(h_2, \dots, h_N) = \frac{p_{T/N}(h_2)p_{T/N}(h_N)}{p_T(1)} \prod_{k=3}^N p_{T/N}(h_k h_{k-1}^{-1}).$$

Consider the equivalent probability measure $\tilde{\mathbb{P}}$ given by

$$d\tilde{\mathbb{P}}/d\mathbb{P} = \rho(H_{l_2}, \dots, H_{l_N})^{-1}.$$

Under $\tilde{\mathbb{P}}$, the random vectors $H_{l_2}E, \dots, H_{l_N}E$ are independent and uniformly distributed on the unit ball in \mathbb{C}^N . Hence

$$\mathbb{P}(\Omega_0) = \tilde{\mathbb{P}}(\Omega_0) = 1.$$

Now, for

$$v = \sum_{k=1}^N a_k \delta_{l_k}$$

we have

$$\langle v, v \rangle_r^N = \left| \sum_{k=1}^N a_k H_{l_k} E \right|^2.$$

So, on the event Ω_0 , $v \in \mathcal{K}_r^N$ only if $a_1 = \dots = a_N = 0$, that is to say, $l_1 + \mathcal{K}_r^N, \dots, l_N + \mathcal{K}_r^N$ are linearly independent in V_r^N . On the other hand, for any loop $l \in \text{Loop}_r(\mathbb{S}_T)$, on the same event Ω_0 , $H_l E$ is a linear combination of $H_{l_1}E, \dots, H_{l_N}E$, so $\delta_l + \mathcal{K}_r^N$ lies in the linear span of $\delta_{l_1} + \mathcal{K}_r^N, \dots, \delta_{l_N} + \mathcal{K}_r^N$. \square

The form $\langle \cdot, \cdot \rangle_x^N$ induces a random Hermitian inner product on V_x^N , which we will denote also by $\langle \cdot, \cdot \rangle_x^N$. Moreover, for all $x, y \in \mathbb{S}_T$ and all $\gamma \in \text{Path}_{x,y}(\mathbb{S}_T)$, we can define a random linear isometry $T_\gamma^N : V_x^N \rightarrow V_y^N$ by

$$T_\gamma^N(v + \mathcal{K}_x^N) = T_\gamma v + \mathcal{K}_y^N.$$

The family of isometries $(T_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ inherits the following properties from $(T_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$:

$$T_x^N = 1_x, \quad T_{\gamma_1 \gamma_2}^N = T_{\gamma_2}^N T_{\gamma_1}^N.$$

Here, we have written x for the constant loop at x , 1_x for the identity map on V_x , and the second identity is valid whenever the concatenation $\gamma_1 \gamma_2$ is possible. On the other hand, if $\gamma_1 \sim \gamma_2$ then $T_{\gamma_1}^N = T_{\gamma_2}^N$. Hence, since $\gamma \gamma^{-1} \sim x$, we have

$$T_{\gamma^{-1}}^N T_\gamma^N = T_{\gamma \gamma^{-1}}^N = T_x^N = 1_x.$$

In fact, the family of isometries $(T_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ may be considered as a Yang–Mills holonomy field in $\text{Isom}(V^N)$, in a sense made precise in the following proposition.

Proposition 2.12 *Conditional on $(V_x^N : x \in \mathbb{S}_T)$, choose a family of independent uniform random isometries $(s(x) : x \in \mathbb{S}_T)$ with $s(x) : \mathbb{C}^N \rightarrow V_x^N$ for all x , and set*

$$\tilde{H}_\gamma^N = s(y)^{-1} T_\gamma^N s(x).$$

Then $(\tilde{H}_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ is a Yang–Mills holonomy field in $U(N)$.

Proof. Consider, for each $x \in \mathbb{S}_T$, the unique linear map $\pi_x^N : \mathcal{V}_x \rightarrow \mathbb{C}^N$ such that $\pi_x^N(\delta_\gamma) = H_\gamma^N E$ for all $\gamma \in \text{Path}_{r,x}(\mathbb{S}_T)$. Then π_N has kernel \mathcal{K}_x^N and the quotient map $V_x^N \rightarrow \mathbb{C}^N$, which we denote abusively also by π_N , is an isometry, by definition of $\langle \cdot, \cdot \rangle_x^N$. For $\gamma_0 \in \text{Path}_{r,x}(\mathbb{S}_T)$ and $\gamma \in \text{Path}_{x,y}(\mathbb{S}_T)$, we have

$$\pi_y^N(T_\gamma \delta_{\gamma_0}) = \pi_y^N(\delta_{\gamma_0 \gamma}) = H_{\gamma_0 \gamma}^N E = H_\gamma^N H_{\gamma_0}^N E = H_\gamma^N \pi_x^N(\delta_{\gamma_0})$$

so the quotient maps satisfy, for all $v \in V_x^N$,

$$\pi_y^N(T_\gamma^N v) = H_\gamma^N \pi_x^N(v).$$

Since Haar measure is invariant under multiplication, the random variables $(\pi_x^N s(x) : x \in \mathbb{S}_T)$ in $U(N)$ are independent, uniformly distributed, and independent of H^N . Now

$$\tilde{H}_\gamma^N = (\pi_y^N s(y))^{-1} H_\gamma^N (\pi_x^N s(x))$$

so $(\tilde{H}_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ is a Yang–Mills holonomy field in $U(N)$ by gauge invariance.

For each $x \in \mathbb{S}_T$, given a path $\gamma \in \text{Path}_{r,x}(\mathbb{S}_T)$, we can define a state τ_γ^N on the set of bounded linear operators on V_x^N by

$$\tau_\gamma^N(A) = \langle [\delta_\gamma]^N, A[\delta_\gamma]^N \rangle_x^N.$$

Then, for all $l \in \text{Loop}_x(\mathbb{S}_T)$,

$$\tau_\gamma^N(T_l^N) = \langle \delta_\gamma, T_l \delta_\gamma \rangle_x^N = \langle \delta_\gamma, \delta_{\gamma l} \rangle_x^N = \tau_T^N(\gamma l^{-1} \gamma^{-1}) = \tau_T^N(l).$$

Then, on restricting τ_γ^N to the von Neumann algebra \mathcal{A}_x^N generated by $(T_l^N : l \in \text{Loop}_x(\mathbb{S}_T))$, we obtain a non-negative unit trace τ_x^N on \mathcal{A}_x^N , which does not depend on the choice of path γ . This construction has been done starting from the random Hermitian forms $\langle \cdot, \cdot \rangle_x^N$ for $x \in \mathbb{S}_T$. We now explore the analogous construction starting from $\langle \cdot, \cdot \rangle_x$.

Consider the kernel

$$\mathcal{K}_x = \{v \in \mathcal{V}_x : \langle v, v \rangle_x = 0\}$$

and write V_x for the Hilbert space obtained by completing $\mathcal{V}_x / \mathcal{K}_x$ with respect to $\langle \cdot, \cdot \rangle_x$. Write $[v] = v + \mathcal{K}_x$. Then $[\delta_{\gamma_1}] = [\delta_{\gamma_2}]$ whenever $\gamma_1 \sim \gamma_2$. For $x, y \in \mathbb{S}_T$ and $\gamma \in \text{Path}_{x,y}(\mathbb{S}_T)$, there is a unique Hilbert space isometry $\tilde{T}_\gamma : V_x \rightarrow V_y$ such that, for all $v \in \mathcal{V}_x$,

$$\tilde{T}_\gamma(v + \mathcal{K}_x) = T_\gamma v + \mathcal{K}_y.$$

The family of isometries $(\tilde{T}_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ then has the following properties:

$$\tilde{T}_x = 1_x, \quad \tilde{T}_{\gamma^{-1}} = (\tilde{T}_\gamma)^{-1}, \quad \tilde{T}_{\gamma_1 \gamma_2} = \tilde{T}_{\gamma_2} \tilde{T}_{\gamma_1}$$

where the last identity holds whenever the concatenation $\gamma_1 \gamma_2$ is possible.

For each $x \in \mathbb{S}_T$, given a path $\gamma \in \text{Path}_{r,x}(\mathbb{S}_T)$, we can define a state τ_γ on the set of bounded linear operators on V_x by

$$\tau_\gamma(A) = \langle [\delta_\gamma], A[\delta_\gamma] \rangle_x.$$

Then, for all $l \in \text{Loop}_x(\mathbb{S}_T)$,

$$\tau_\gamma(\tilde{T}_l) = \langle \delta_\gamma, T_l \delta_\gamma \rangle_x = \langle \delta_\gamma, \delta_{\gamma l} \rangle_x = \Phi_T(\gamma l^{-1} \gamma^{-1}) = \Phi_T(l).$$

Recall from Proposition 2.3 that $\Phi_T(x) = 1$ and $\Phi_T(l_1 l_2) = \Phi_T(l_2 l_1)$. Then, on restricting τ_γ to the von Neumann algebra \mathcal{A}_x generated by $(\tilde{T}_l : l \in \text{Loop}_x(\mathbb{S}_T))$, we obtain a non-negative unit trace τ_x on \mathcal{A}_x , which does not depend on the choice of path γ .

We note some further properties of (\mathcal{A}_x, τ_x) . First, for all integers n , and all $l \in \text{Loop}_x(\mathbb{S}_T)$,

$$\tau_x(\tilde{T}_l^n) = \Phi_T(l^n) = \int_{\mathbb{U}} \omega^n \nu_T(l)(d\omega)$$

where $\nu_T(l)$ is the limit spectral measure obtained in Subsection 2.6. So $\nu_T(l)$ is the spectral measure of \tilde{T}_l . (Here T refers to the area of the sphere, while \tilde{T}_l^n is the n th power of the transport operator \tilde{T}_l defined above.) Second, since the master field is invariant under area-preserving diffeomorphisms of \mathbb{S}_T , the choice of such a diffeomorphism ψ gives an isomorphism $(\mathcal{A}_x, \tau_x) \rightarrow (\mathcal{A}_y, \tau_y)$ whenever $\psi(x) = y$.

Singer [52] conjectured, without explicit construction, that the von Neumann algebras \mathcal{A}_x were factors, that is to say, their centres were trivial³. We remark that, if this conjecture holds then, since⁴ the spectral measures $\nu_T(l)$ are absolutely continuous, at least for simple loops separating the sphere into components of equal area, as follows from Proposition 6.2, and since τ_x is a finite normalized trace, we see that

$$\{\tau_x(p) : p \in \mathcal{A}_x, p^2 = p\} = [0, 1]$$

and \mathcal{A}_x must be of type II_1 and have unique state τ_x .

It is an open question whether in fact \mathcal{K}_x is spanned, for all x , by vectors of the form $\delta_\gamma - \delta_{\gamma_0}$, where $\gamma, \gamma_0 \in \text{Path}_{r,x}(\mathbb{S}_T)$ and, for some sequence $(\gamma_n : n \in \mathbb{N})$ in $\text{Path}_{r,x}(\mathbb{S}_T)$ with $\gamma_n \sim \gamma_0$ for all n , we have $\gamma_n \rightarrow \gamma$ in length. Since we know that such vectors all lie in \mathcal{K}_x^N for all N , if true, this would allow to identify V_x^N with the orthogonal complement of $\mathcal{K}_x^N / \mathcal{K}_x$ in V_x . Then the Yang–Mills holonomy field $(T_\gamma^N : \gamma \in \text{Path}(\mathbb{S}_T))$ in $\text{Isom}(V^N)$ would be obtained by restricting the family of isometries $(\tilde{T}_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ to the N -dimensional random subspaces $(V_x^N : x \in \mathbb{S}_T)$.

3. Harmonic analysis in $U(N)$ and a discrete β -ensemble

3.1. A representation formula. Let $H = (H_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ be a Yang–Mills holonomy field in $U(N)$. Here, and from now on, we suppress mention of N in the notation for H . On the other hand, for $l \in \text{Loop}(\mathbb{S}_T)$ and $n \in \mathbb{Z}$, we will write H_l^n for the n th power of the matrix H_l which, by the multiplicative property, is also given by H_{l^n} . We obtain in this subsection a key formula for the moments of the holonomy H_l of a simple loop l in terms of a certain discrete β -ensemble, with $\beta = 2$. Set

$$\mathbb{Z}_{\text{sym}} = \begin{cases} \mathbb{Z}, & \text{if } N \text{ is odd,} \\ \mathbb{Z} + 1/2, & \text{if } N \text{ is even.} \end{cases}$$

Consider the discrete β -ensemble Λ in $N^{-1}\mathbb{Z}_{\text{sym}}$ given by

$$\mathbb{P}(\Lambda = \lambda) \propto \prod_{\substack{j,k=1 \\ j < k}}^N (\lambda_j - \lambda_k)^2 \prod_{i=1}^N e^{-N\lambda_i^2 T/2} \quad (16)$$

where λ runs over decreasing sequences $(\lambda_1, \dots, \lambda_N)$ in $N^{-1}\mathbb{Z}_{\text{sym}}$. For $\alpha \in \mathbb{R} \setminus \{0\}$ and for $z \in \mathbb{C}$ with $|\alpha||z - \lambda_j| > 1$ for all j , set

$$G_\lambda^\alpha(z) = \frac{\alpha}{N} \sum_{j=1}^N \text{Log} \left(1 + \frac{1}{\alpha(z - \lambda_j)} \right)$$

³ See for example [53].

⁴ See for example, Section 8.4 of [34].

where Log denotes the principal value of the logarithm. Then, for $a \in (0, T)$, set $I_0^a(\lambda) = 1$ and define for $n \in \mathbb{Z} \setminus \{0\}$

$$I_n^a(\lambda) = \frac{e^{-an^2/(2N)}}{2\pi in} \int_{\gamma} \exp\{-n(az - G_{\lambda}^{N/n}(z))\} dz$$

where γ is any positively oriented simple loop around the set

$$[\lambda_N, \lambda_1] + \{z \in \mathbb{C} : |z| \leq |n|/N\}.$$

It is straightforward to check that $I_n^a(\lambda)$ does not depend on the choice of γ .

Proposition 3.1 *Let $l \in \text{Loop}(\mathbb{S}_T)$ be a simple loop which divides \mathbb{S}_T into components of areas a and b . Then, for all $m, n \in \mathbb{Z}$,*

$$\mathbb{E}(\text{tr}(H_l^{-m}) \text{tr}(H_l^n)) = \mathbb{E}(I_m^a(\Lambda) I_n^b(\Lambda)).$$

Here and from now on, we suppress the N in our notation for the normalized trace on $U(N)$. This formula allows to prove the convergence of the random variables $\text{tr}(H_l^n)$ for simple loops l , as will be explained in Subsection 3.2.

To prove Proposition 3.1, we will use the decomposition of the heat kernel as a sum over the characters of $U(N)$. The results we use may be found for example in [37]. For $\lambda \in (\mathbb{Z}_{\text{sym}})^N$, set

$$\|\lambda\|^2 = \frac{1}{N} \sum_{j=1}^N \lambda_j^2.$$

Write $\rho = (\rho_1, \dots, \rho_N)$ for the unique minimizer of $\|\cdot\|$ among decreasing sequences in $(\mathbb{Z}_{\text{sym}})^N$, which is given by

$$\rho_j = \frac{1}{2}(N+1) - j.$$

For $\lambda \in \mathbb{Z}^N$, there is a unique continuous function $\chi_{\lambda} : U(N) \rightarrow \mathbb{C}$ given by the Weyl character formula

$$\chi_{\lambda}(g) \det(e^{i\theta_j \rho_k})_{j,k=1}^N = \det(e^{i\theta_j (\lambda_k + \rho_k)})_{j,k=1}^N, \quad g \in U(N) \quad (17)$$

where $e^{i\theta_1}, \dots, e^{i\theta_N}$ are the eigenvalues of g . Then

$$(\chi_{\lambda} : \lambda \in \mathbb{Z}^N, \lambda_1 \geq \dots \geq \lambda_N)$$

is a parametrization of the set of characters of irreducible representations of $U(N)$. For characters χ_{λ} and χ_{μ} , we have

$$\int_{U(N)} \chi_{\lambda}(g) \overline{\chi_{\mu}(g)} dg = \int_{U(N)} \chi_{\lambda}(g) \chi_{\mu}(g^{-1}) dg = \delta_{\lambda, \mu} \quad (18)$$

and

$$\Delta \chi_{\lambda} = -(\|\lambda + \rho\|^2 - \|\rho\|^2) \chi_{\lambda}. \quad (19)$$

Moreover, the heat kernel $(p_t(g) : t \in (0, \infty), g \in U(N))$ is given by the following absolutely convergent sum

$$p_t(g) = e^{\|\rho\|^2 t/2} \sum_{\lambda} \chi_{\lambda}(1) \chi_{\lambda}(g) e^{-\|\lambda + \rho\|^2 t/2}. \quad (20)$$

The character values at the identity are given by the Weyl dimension formula

$$\chi_{\lambda}(1) = \prod_{\substack{j,k=1 \\ j < k}}^N \frac{\lambda_j + \rho_j - \lambda_k - \rho_k}{\rho_j - \rho_k}. \quad (21)$$

The change of variable $\mu = \lambda + \rho$ gives a convenient reparametrization of the set of characters by

$$W = \{\mu \in (\mathbb{Z}_{\text{sym}})^N : \mu_1 > \cdots > \mu_N\}.$$

For $x \in (\mathbb{Z}_{\text{sym}})^N$ with all components distinct, we will write $[x]$ for the decreasing rearrangement of x . From (17), we see that,

$$\chi_{x-\rho} = \varepsilon(x)\chi_{[x]-\rho}$$

where

$$\varepsilon(x) = \begin{cases} \text{sgn}(\sigma), & \text{if } x \text{ has all components distinct,} \\ 0, & \text{otherwise,} \end{cases}$$

where σ is the unique permutation such that $[x]_j = x_{\sigma(j)}$ for all j . Then the orthogonality relation (18) extends to all $x, y \in (\mathbb{Z}_{\text{sym}})^N$ in the form

$$\int_{U(N)} \chi_{x-\rho}(g) \chi_{y-\rho}(g^{-1}) dg = \varepsilon(x) \varepsilon(y) \delta_{[x],[y]}. \quad (22)$$

To compute the desired moments of holonomy traces, we shall need the following product formula, which may be obtained from (17) by a straightforward computation. For all $n \in \mathbb{Z}$, we have

$$\chi_\lambda(g) \text{Tr}(g^n) = \sum_{j=1}^N \chi_{\lambda+n\omega^j}(g). \quad (23)$$

where ω^j is the j th elementary vector in \mathbb{Z}^N .

Proof of Proposition 3.1. From the definition of the Yang–Mills measure, we have

$$\mathbb{E}(\text{tr}(H_l^{-m}) \text{tr}(H_l^n)) \propto \int_{U(N)} p_a(g) \text{tr}(g^{-m}) \text{tr}(g^n) p_b(g^{-1}) dg$$

where \propto signifies equality up to a constant independent of m and n . We expand the heat kernel in characters to obtain

$$\begin{aligned} & \int_{U(N)} p_a(g) \text{tr}(g^{-m}) \text{tr}(g^n) p_b(g^{-1}) dg \\ & \propto \sum_{\lambda, \mu \in W} e^{-\|\lambda\|^2 a/2 - \|\mu\|^2 b/2} \chi_{\lambda-\rho}(1) \chi_{\mu-\rho}(1) \int_{U(N)} \chi_{\lambda-\rho}(g) \text{tr}(g^{-m}) \text{tr}(g^n) \chi_{\mu-\rho}(g^{-1}) dg. \end{aligned}$$

The interchange of summation and integration here is valid because $a, b > 0$ which ensures absolute convergence. By orthogonality of characters (22) and the product rule (23), for all $\lambda, \mu \in W$,

$$\begin{aligned} & \int_{U(N)} \chi_{\lambda-\rho}(g) \text{tr}(g^{-m}) \text{tr}(g^n) \chi_{\mu-\rho}(g^{-1}) dg \\ & = \frac{1}{N^2} \sum_{j,k=1}^N \varepsilon(\lambda - m\omega^j) \varepsilon(\mu - n\omega^k) \delta_{[\lambda-m\omega^j],[\mu-n\omega^k]}. \end{aligned}$$

Now, for $\nu \in W$, we have $[\lambda - m\omega^j] = [\mu - n\omega^k] = \nu$ for some $j, k \in \{1, \dots, N\}$ if and only if $\lambda = [\nu + m\omega^{j'}]$ and $\mu = [\nu + n\omega^{k'}]$ for some $j', k' \in \{1, \dots, N\}$, and then

$$N\|\lambda\|^2 = N\|\nu\|^2 + 2m\nu_{j'} + m^2, \quad N\|\mu\|^2 = N\|\nu\|^2 + 2n\nu_{k'} + n^2$$

and

$$\varepsilon(\lambda - m\omega^j) = \varepsilon(\nu + m\omega^{j'}), \quad \varepsilon(\mu - m\omega^k) = \varepsilon(\nu + n\omega^{k'})$$

so, using the dimension formula (21),

$$\begin{aligned} \chi_{\lambda-\rho}(1)\varepsilon(\lambda - m\omega^j) &= \chi_{\nu+m\omega^{j'}-\rho}(1) = \chi_{\nu-\rho}(1) \prod_{i \neq j} \frac{\nu_j + m - \nu_i}{\nu_j - \nu_i}, \\ \chi_{\mu-\rho}(1)\varepsilon(\mu - m\omega^k) &= \chi_{\nu+n\omega^{k'}-\rho}(1) = \chi_{\nu-\rho}(1) \prod_{i \neq k} \frac{\nu_k + n - \nu_i}{\nu_k - \nu_i}. \end{aligned}$$

Hence

$$\begin{aligned} &\int_{U(N)} p_a(g) \operatorname{tr}(g^{-m}) \operatorname{tr}(g^n) p_b(g^{-1}) dg \\ &\propto \sum_{\nu \in W} \prod_{\substack{j,k=1 \\ j < k}}^N (\nu_j - \nu_k)^2 e^{-\|\nu\|^2 T/2} J(\nu, m, a) J(\nu, n, b) \end{aligned}$$

where

$$J(\nu, m, a) = e^{-m^2 a/(2N)} \frac{1}{N} \sum_{j=1}^N e^{-m a \nu_j / N} \prod_{i \neq j} \frac{\nu_j + m - \nu_i}{\nu_j - \nu_i}.$$

Note that $J(\nu, 0, a) = 1 = I_0^a(\nu/N)$ and, for $|m| \geq 1$,

$$\begin{aligned} J(\nu, m, a) &= \frac{e^{-m^2 a/(2N)}}{2\pi i m N} \int_{N\gamma(\nu)} \prod_{j=1}^N \left(1 + \frac{m}{z - \nu_j}\right) e^{-m a z / N} dz \\ &= \frac{e^{-m^2 a/(2N)}}{2\pi i m} \int_{\gamma(\nu)} \exp\{-m(a z - G_{\nu/N}^{N/m}(z))\} dz = I_m^a(\nu/N) \end{aligned}$$

where $\gamma(\nu)$ is a positively oriented simple loop around

$$[\nu_N/N, \nu_1/N] + \{z \in \mathbb{C} : |z| \leq |m|/N\}.$$

So we obtain

$$\begin{aligned} \mathbb{E}(\operatorname{tr}(H_l^{-m}) \operatorname{tr}(H_l^n)) &\propto \sum_{\nu \in W} \prod_{\substack{j,k=1 \\ j < k}}^N (\nu_j - \nu_k)^2 e^{-\|\nu\|^2 T/2} I_m^a(\nu/N) I_n^b(\nu/N) \\ &\propto \sum_{N\lambda \in W} \prod_{\substack{j,k=1 \\ j < k}}^N (\lambda_j - \lambda_k)^2 \prod_{i=1}^N e^{-N\lambda_i^2 T/2} I_m^a(\lambda) I_n^b(\lambda) \\ &\propto \mathbb{E}(I_m^a(\Lambda) I_n^b(\Lambda)). \end{aligned}$$

Since the identity $\mathbb{E}(\operatorname{tr}(H_l^{-m}) \operatorname{tr}(H_l^n)) = \mathbb{E}(I_m^a(\Lambda) I_n^b(\Lambda))$ holds for $m = n = 0$, it therefore holds for all m and n . \square

The first part of the above proof follows ideas from the physics literature [6, 14]. The use of contour integrals in writing the function J and in the formulation of Proposition 3.1 is new and provides us with a route to make rigorous the asymptotics performed in [6, 14].

3.2. Concentration for the discrete β -ensemble and tightness of the support. We shall need two facts about the discrete β -ensemble Λ defined in equation (16). Denote by π_N the law on $\mathcal{M}_1(\mathbb{R})$ of the normalized empirical distribution

$$\mu_\Lambda = \frac{1}{N} \sum_{i=1}^N \delta_{\Lambda_i}.$$

Recall from (5) the functional

$$\mathcal{I}_T(\mu) = \int_{\mathbb{R}^2} \left\{ \frac{1}{2}(x^2 + y^2)T - 2 \log |x - y| \right\} \mu(dx) \mu(dy)$$

defined for probability measures μ on \mathbb{R} such that $\mu([a, b]) \leq b - a$ for all intervals $[a, b]$. We extend \mathcal{I}_T to $\mathcal{M}_1(\mathbb{R})$ by setting $\mathcal{I}_T(\mu) = \infty$ if μ does not satisfy this constraint. Guionnet and Maïda [30] showed the following large deviation principle.

Theorem 3.2 *The sequence of probability measures $(\pi_N : N \in \mathbb{N})$ satisfies a large deviation principle on $\mathcal{M}_1(\mathbb{R})$ with rate function \mathcal{I}_T and speed N^2 .*

Let us remark that this result also allows to prove, for all $T \in (0, \infty)$, the existence of the limit

$$F(T) = \lim_{N \rightarrow \infty} N^{-2} \log(p_T^N(1)) \quad (24)$$

where $p_T^N(1)$ denotes the heat kernel of $U(N)$ on the diagonal at time T . This approach was followed by Lévy and Maïda, who obtained in [44, Proposition 5.2] an exact formula for F . They showed moreover that F is C^2 on $(0, \infty)$ and C^∞ on $(0, \pi^2) \cup (\pi^2, \infty)$, but that the third derivative has a discontinuity at π^2 . In doing so, they gave a rigorous proof of the *Douglas–Kazakov phase transition* [15] and of the fact that it is of third order. See also [7] for another approach using tools of statistical mechanics. We call $T \in (0, \pi^2)$ the *subcritical regime* and $T \in (\pi^2, \infty)$ the *supercritical regime*. We shall see in Subsections 6.1 and 6.2 that, in the limit $N \rightarrow \infty$, the behaviour of the eigenvalues of the unitary Brownian loop of length T is very different in one regime to the other.

We need also a tightness result for the positions Λ_N and Λ_1 of the leftmost and rightmost particles, which is obtained by a variation on ideas of Johansson [32]. See also Féral [20], who adapts to the discrete case some arguments of Ben Arous, Dembo and Guionnet [3, Section 6] for eigenvalues of GOE matrices.

Lemma 3.3 *Set*

$$\Lambda^* = \max\{|\Lambda_1|, |\Lambda_N|\}.$$

For all $a \in [0, \infty)$, there are constants $C, R < \infty$ depending only on a and T such that

$$\mathbb{E} \left(e^{a\Lambda^*} 1_{\{\Lambda^* \geq R\}} \right) \leq C e^{-N}.$$

Proof. It will be convenient in this proof to label the particle positions in increasing order, where before we labelled them in decreasing order, so Λ_N now denotes the position of the rightmost particle. Then, by symmetry, it will suffice to show that, for all $a \in [0, \infty)$, there are constants $C, R < \infty$ depending only on a and T such that

$$\mathbb{E} \left(e^{a\Lambda_N} 1_{\{\Lambda^* = \Lambda_N \geq R\}} \right) \leq C e^{-N}.$$

Fix N and, for $M \in \mathbb{N}$, set

$$Z_M = \sum_{\lambda} \prod_{\substack{j,k=1 \\ j < k}}^M (\lambda_j - \lambda_k)^2 \prod_{i=1}^M e^{-N\lambda_i^2 T/2}$$

where the sum is taken over the set S_M of increasing sequences $\lambda = (\lambda_1, \dots, \lambda_M)$ in $N^{-1}\mathbb{Z}_{\text{sym}}$. Only the cases $M = N-1$ and $M = N$ will be considered further. In the following calculation, we write the possible values of $\Lambda = (\Lambda_1, \dots, \Lambda_N)$ in the form (λ, λ_N) , where $\lambda = (\lambda_1, \dots, \lambda_{N-1}) \in S_{N-1}$ and $\lambda_N \in N^{-1}\mathbb{Z}_{\text{sym}}$, and we write λ^* for $\max\{|\lambda_1|, |\lambda_{N-1}|\}$. We have

$$\begin{aligned} & \mathbb{E} \left(e^{a\Lambda_N} 1_{\{\Lambda^* = \Lambda_N \geq R\}} \right) \\ &= \frac{1}{Z_N} \sum_{\lambda_N} \sum_{\lambda} e^{a\lambda_N} 1_{\{\lambda_N \geq R, \lambda^* \leq \lambda_N\}} \prod_{\substack{j,k=1 \\ j < k}}^N (\lambda_j - \lambda_k)^2 \prod_{i=1}^N e^{-N\lambda_i^2 T/2} \\ &\leq \frac{1}{Z_N} \sum_s \sum_{\lambda} e^{as} 1_{\{s \geq R, \lambda^* \leq s\}} \prod_{\substack{j,k=1 \\ j < k}}^{N-1} (\lambda_j - \lambda_k)^2 \prod_{i=1}^{N-1} (s - \lambda_i)^2 \prod_{i=1}^{N-1} e^{-N\lambda_i^2 T/2} e^{-Ns^2 T/2} \\ &\leq \frac{Z_{N-1}}{Z_N} \sum_s e^{as - Ns^2 T/2} 1_{\{s \geq R\}} (4s^2)^{N-1} \end{aligned}$$

where s and λ_N are summed over $N^{-1}\mathbb{Z}_{\text{sym}}$ and λ is summed over S_{N-1} . We will show in Lemma 3.4 that there is a constant $c \in (0, \infty)$, depending only on T , such that

$$Z_{N-1}/Z_N \leq e^{cN}.$$

We can now choose $C, R \in (0, \infty)$, depending only on a, c and T , so that

$$\sum_s e^{as - Ns^2 T/2} 1_{\{s \geq R\}} (4s^2)^{N-1} \leq C e^{-(c+1)N}$$

for all N , so obtaining the desired estimate

$$\mathbb{E} \left(e^{a\Lambda_N} 1_{\{\Lambda^* = \Lambda_N \geq R\}} \right) \leq C e^{-N}.$$

□

It remains to prove the following estimate, which limits the rate of decay of Z_N as $N \rightarrow \infty$.

Lemma 3.4 *There exists $c \in (0, \infty)$, depending only on T , such that, for all $N \geq 2$,*

$$Z_{N-1}/Z_N \leq e^{cN}.$$

Proof. Let us consider again the set S_{N-1} of increasing sequences $\lambda = (\lambda_1, \dots, \lambda_{N-1})$ in $N^{-1}\mathbb{Z}_{\text{sym}}$. For $\lambda \in S_{N-1}$, set

$$E(\lambda) = \prod_{j < k} (\lambda_j - \lambda_k)^2 \prod_i e^{-(N-1)T\lambda_i^2/2}$$

and set

$$A_N = \frac{1}{N} \sum_{s \in N^{-1}\mathbb{Z}_{\text{sym}}} e^{-Ts^2/2}, \quad E_N = \sup_{\lambda \in S_{N-1}} E(\lambda).$$

Note that

$$A_N \leq \frac{1}{N} + \int_{\mathbb{R}} e^{-Tx^2/2} dx \leq 1 + \sqrt{\frac{2\pi}{T}}.$$

We will show that, for $r = 1 + 4/T$, there exists $\lambda(N) \in S_{N-1}$ with $\lambda(N)^* \leq r$ such that

$$E_N = E(\lambda(N)). \tag{25}$$

Now

$$\begin{aligned} Z_{N-1} &= \sum_{\lambda \in S_{N-1}} \prod_{j < k} (\lambda_j - \lambda_k)^2 \prod_i e^{-NT\lambda_i^2/2} = \sum_{\lambda \in S_{N-1}} E(\lambda) \prod_i e^{-T\lambda_i^2/2} \\ &\leq \sum_{\lambda \in S_{N-1}} E_N \prod_i e^{-T\lambda_i^2/2} = (N^N/N!) E_N A_N^{N-1} \leq e^N E_N A_N^{N-1}. \end{aligned}$$

On the other hand, there exists $s \in N^{-1}\mathbb{Z}_{\text{sym}}$ with $s \in [2r, 3r]$ so, by considering the single term in the sum where $\lambda = (\lambda(N), s)$,

$$\begin{aligned} Z_N &= \sum_{\lambda \in S_N} \prod_{j < k} (\lambda_j - \lambda_k)^2 \prod_i e^{-NT\lambda_i^2/2} \\ &\geq \prod_{j < k} (\lambda_j(N) - \lambda_k(N))^2 \prod_i (\lambda_i(N) - s)^2 \prod_i e^{-NT\lambda_i(N)^2/2} e^{-NTs^2/2} \\ &= E_N \prod_i (\lambda_i(N) - s)^2 \prod_i e^{-T\lambda_i(N)^2/2} e^{-NTs^2/2} \\ &\geq E_N e^{-(N-1)Tr^2/2} e^{-9NTs^2/2}. \end{aligned}$$

Hence

$$Z_{N-1}/Z_N \leq e^N A_N^{N-1} e^{(N-1)Tr^2/2} e^{9NTs^2/2}$$

which is a bound of the desired form.

It remains to show (25). To see this, given $\lambda \in S_{N-1}$ with $\lambda_{N-1} = \lambda^* = t \geq 1 + 4/T$, we can choose $s \in N^{-1}\mathbb{Z}_{\text{sym}} \setminus \{\lambda_1, \dots, \lambda_{N-2}\}$ with $|s| \leq 1$ and consider the increasing rearrangement $\tilde{\lambda}$ of $(\lambda_1, \dots, \lambda_{N-2}, s)$. Note that

$$\sum_{i=1}^{N-2} \log |s - \lambda_i| \geq \sum_{x \in N^{-1}\mathbb{Z}_{\text{sym}} \setminus \{s\} : |x-s| \leq 1} \log |s - x| \geq 2N \int_0^1 \log x dx = -2N$$

whereas

$$T(t^2 - 1)/2 - 2 \log t - 4 \geq T(t^2 - 1)/2 - 2t - 4 \geq 0.$$

Then

$$E(\lambda)/E(\tilde{\lambda}) \leq t^{2N} \prod_{i=1}^{N-2} (s - \lambda_i)^{-2} e^{-NT(t^2-1)/2} \leq \exp\{-N(T(t^2 - 1)/2 - 2 \log t - 4)\} \leq 1.$$

A similar argument applies if $\lambda_1 = -\lambda^* \leq -1 - 4/T$. By iterating this procedure, we can find $\mu \in S_{N-1}$ with $\mu^* \leq 1 + 4/T$ and $E(\mu) \geq E(\lambda)$. Since there are only finitely many sequences $\mu \in S_{N-1}$ with $\mu^* \leq 1 + 4/T$, this establishes the claim. \square

3.3. Dimension-free continuity estimate for the holonomy of a simple loop. The following estimate will be needed for the proof of Proposition 2.7.

Lemma 3.5 *There is a universal constant $K_T \in (0, \infty)$, in particular independent of N , such that, for any simple loop $l \in \text{Loop}(\mathbb{S}_T)$ dividing \mathbb{S}_T into components of areas a and b ,*

$$\mathbb{E}(\text{tr}(I - H_l)) \leq K \min(a, b).$$

From the following proof, it should be possible to show that K_T is bounded as a function of T . We shall not use this fact and will not prove it here.

Proof. By symmetry, it suffices to consider the case $b \leq a$. Define, for each decreasing sequence $\lambda = (\lambda_1, \dots, \lambda_N)$ in $N^{-1}\mathbb{Z}_{\text{sym}}$,

$$D(b, \lambda) = \frac{1}{2\pi i} \int_{\gamma} e^{-b(z+1/(2N))} \exp\{G_{\lambda}^N(z)\} dz$$

where γ is a positively oriented simple loop around the set $[\lambda_N, \lambda_1] + \{z \in \mathbb{C} : |z| \leq 1/N\}$. We use the residue theorem to compute

$$D(b, \lambda) = \frac{1}{N} \sum_{j=1}^N e^{-b(\lambda_j+1/(2N))} \prod_{i \neq j} \frac{\lambda_j + N^{-1} - \lambda_i}{\lambda_j - \lambda_i}.$$

Note the identity

$$\frac{1}{N} \sum_{j=1}^N \prod_{i \neq j} \frac{\lambda_j + N^{-1} - \lambda_i}{\lambda_j - \lambda_i} = 1. \quad (26)$$

This may be seen, for example, by evaluating both sides of the product rule (23) for $\chi_{\rho+N\lambda}(g)\text{Tr}(g)$ at $g = 1$ using the dimension formula (21). Moreover, since $\lambda \in N^{-1}\mathbb{Z}_{\text{sym}}$, all the terms in the sum (26) are non-negative. Now, for all j ,

$$1 - e^{-b(\lambda_j+1/(2N))} \leq b|\lambda_j + 1/(2N)| \leq b(\lambda^* + 1/(2N))$$

so

$$1 - D(b, \lambda) = \frac{1}{N} \sum_{j=1}^N (1 - e^{-b(\lambda_j+1/(2N))}) \prod_{i \neq j} \frac{\lambda_j + N^{-1} - \lambda_i}{\lambda_j - \lambda_i} \leq b(\lambda^* + 1/(2N)).$$

But, by Proposition 3.1, we have

$$\mathbb{E}(\text{tr}(H_l)) = \mathbb{E}(D(b, \Lambda))$$

so

$$\mathbb{E}(\text{tr}(I - H_l)) = \mathbb{E}(1 - D(b, \Lambda)) \leq b \left(\mathbb{E}(\Lambda^*) + \frac{1}{2N} \right)$$

as required. \square

3.4. Evaluation of some contour integrals. In passing from the limit particle density ρ_T for the β -ensemble, as given in (8), to the evaluation of the master field on simple loops, we will need to evaluate certain contour integrals expressed in terms of the Stieltjes transform

$$G_T(z) = \int_{\mathbb{R}} \frac{\rho_T(x) dx}{z - x}.$$

The following calculation is taken from [6, 14].

Proposition 3.6 *Let $T \in (0, \pi^2]$ and let $a \in (0, T)$. Let γ be a positively oriented closed curve around the set $[-2/\sqrt{T}, 2/\sqrt{T}]$. Then, for all $n \in (0, \infty)$,*

$$\frac{1}{2\pi i n} \int_{\gamma} \exp\{-n(az - G_T(z))\} dz = \int_{\mathbb{R}} e^{inx} s_{a(T-a)/T}(x) dx$$

where s_t is the semi-circle density (15) of variance t .

Proof. Since $T \in (0, \pi^2]$, we have $\rho_T = s_{1/T}$. Then $\rho_T(x) = \sqrt{T}\rho_1(\sqrt{T}x)$ so, by a scaling argument, it will suffice to consider the case $T = 1$. A standard calculation of the Stieltjes transform gives

$$G_1(z) = \int_{\mathbb{R}} \frac{\rho_1(x)dx}{z-x} = \frac{z - \sqrt{z^2 - 4}}{2}.$$

Note that G_1 maps $\mathbb{C} \setminus [-2, 2]$ conformally to the punctured unit disc $\mathbb{D} \setminus \{0\}$ with inverse $z + 1/z$. Also, $G_1(\gamma)$ is a negatively oriented closed curve around $\{0\}$. Write $b = 1 - a$. We make the change of variable $w = G_1(z)$ to obtain

$$\begin{aligned} \frac{1}{2\pi i n} \int_{\gamma} \exp\{-n(az - G_1(z))\} dz &= \frac{1}{2\pi i n} \int_{G_1(\gamma)} \exp\{n(bw - aw^{-1})\} (1 - w^{-2}) dw \\ &= \frac{1}{2\pi n} \int_0^{2\pi} \exp\{n(be^{-i\theta} - ae^{i\theta})\} (e^{i\theta} - e^{-i\theta}) d\theta \\ &= \frac{1}{2\pi n} \sum_{k=0}^{\infty} \frac{n^k}{k!} \sum_{j=0}^k \binom{k}{j} \int_0^{2\pi} (be^{-i\theta})^j (-ae^{i\theta})^{k-j} (e^{i\theta} - e^{-i\theta}) d\theta \\ &= \sum_{m=0}^{\infty} \frac{(-n^2 ab)^m}{m!(m+1)!} = \int_{\mathbb{R}} e^{inx} s_{ab}(x) dx \end{aligned}$$

where we used in the last equality the moment formula

$$\int_{\mathbb{R}} x^{2m} s_t(x) dx = \frac{t^{2m}}{m!(m+1)!}.$$

□

More generally, for all $T \in (0, \infty)$, the following is obtained in [44, equation (4.12)]

$$G_T(z) = \frac{zT}{2} - \frac{2}{\beta z} \sqrt{(z^2 - \alpha^2)(z^2 - \beta^2)} \int_0^1 \frac{ds}{(1 - \alpha^2 s^2/z^2) \sqrt{(1 - s^2)(1 - k^2 s^2)}} \quad (27)$$

where $k = \alpha/\beta \in (0, 1)$ and α, β are as defined in (7). Moreover, for $|x| \in [\alpha, \beta]$, in the limit $z \rightarrow x$ with $z \notin \mathbb{R}$, we have

$$\operatorname{Re}(G_T(z)) \rightarrow \lim_{\varepsilon \rightarrow 0} \int_{|y-x| \geq \varepsilon} \frac{\rho_T(y) dy}{x-y} = \frac{xT}{2}. \quad (28)$$

Proposition 3.7 *Let $T \in (0, \infty)$ and let $a, b \in (0, T)$ with $a + b = T$. Let γ be a positively oriented closed curve around the set $[-\beta, \beta]$. Then, for all $n \in \mathbb{N}$,*

$$\begin{aligned} \frac{1}{2\pi i n} \int_{\gamma} \exp\{-n(az - G_T(z))\} dz &= \frac{2}{n\pi} \int_0^{\infty} \cosh\{(a-b)nx/2\} \sin\{n\pi\rho_T(x)\} dx \\ &= \frac{1}{2\pi i n} \int_{\gamma^{-1}} \exp\{n(bz - G_T(z))\} dz. \end{aligned}$$

Proof. Since the integrand of the left-hand side is holomorphic in $\mathbb{C} \setminus [-\beta, \beta]$, we can take γ to be the anti-clockwise boundary of $[-\beta - \varepsilon, \beta + \varepsilon] \times [-\varepsilon, \varepsilon]$ for any $\varepsilon > 0$. Now, as ρ_T is Hölder continuous, by the Plemelj–Sokhotskyi formula [25], G_T can be continuously extended, as G_+ and G_- say, on $\overline{\mathbb{H}} = \{z \in \mathbb{C} : \operatorname{Im}(z) \geq 0\}$ and $-\overline{\mathbb{H}}$, with

$$G_{\pm}(x) = \lim_{\varepsilon \rightarrow 0} \int_{|y-x| \geq \varepsilon} \frac{\rho_T(y) dy}{x-y} \mp i\pi\rho_T(x) = \frac{xT}{2} \mp i\pi\rho_T(x)$$

for any $x \in \mathbb{R}$. We can take the limit $\varepsilon \rightarrow 0$ in the contour integrals along γ and γ^{-1} , using the dominated convergence theorem, to obtain

$$\frac{1}{n\pi} \int_{\mathbb{R}} \exp\{(a-b)nx/2\} \sin\{n\pi\rho_T(x)\} dx.$$

Since ρ_T is symmetric, this gives the claimed identity. \square

3.5. Proof of Proposition 2.5. Consider the discrete β -ensemble Λ defined by (16). By Theorem 3.2,

$$\mu_\Lambda \rightarrow \mu_T \quad \text{weakly in probability on } \mathbb{R} \text{ as } N \rightarrow \infty. \quad (29)$$

Fix $n \in \mathbb{N}$. By Lemma 3.3, there exist $C, R \in (0, \infty)$, independent of N , such that

$$\mathbb{E}(e^{2nT\Lambda^*} 1_{\Omega_R^c}) \leq Ce^{-N} \quad (30)$$

where

$$\Lambda^* = \max\{|\Lambda_1|, |\Lambda_N|\}, \quad \Omega_R = \{\text{supp}(\mu_\Lambda) \subseteq [-R, R]\} = \{\Lambda^* \leq R\}.$$

We increase the value of R if necessary so that

$$\text{supp}(\mu_T) \subseteq [-R, R].$$

Denote by γ_R the positively oriented boundary of the set

$$[-R, R] + \{z \in \mathbb{C} : |z| \leq 1\}.$$

Recall that, for $\alpha \in (0, \infty)$ and $\text{dist}(z, \text{supp}(\mu_\Lambda)) > 1/\alpha$, we set

$$G_\Lambda^\alpha(z) = \alpha \int_{\mathbb{R}} \text{Log} \left(1 + \frac{1}{\alpha(z-x)} \right) \mu_\Lambda(dx).$$

For $N \geq n+1$, the contour $\gamma_{R \vee \Lambda^*}$ contains the set

$$\text{supp}(\mu_\Lambda) + \{z \in \mathbb{C} : |z| \leq n/N\}$$

so we can write, for $a \in (0, T)$,

$$I_n^a(\Lambda) = \frac{e^{-an^2/(2N)}}{2\pi in} \int_{\gamma_{R \vee \Lambda^*}} \exp\{-n(az - G_\Lambda^{N/n}(z))\} dz.$$

Recall also that we set

$$G_T(z) = \int_{\mathbb{R}} \frac{\mu_T(dx)}{z-x}$$

and, for $a, b > 0$ with $a+b=T$,

$$I_n^a = I_n^b = \frac{2}{n\pi} \int_0^\infty \cosh\{(a-b)nx/2\} \sin\{n\pi\rho_T(x)\} dx$$

and that, by Proposition 3.7,

$$I_n^a = \frac{1}{2\pi in} \int_{\gamma_R} \exp\{-n(az - G_T(z))\} dz.$$

In Proposition 3.1 we showed that, for any simple loop $l \in \text{Loop}(\mathbb{S}_T)$ which divides \mathbb{S}_T into components of areas a and b ,

$$\mathbb{E}(\text{tr}(H_l^n)) = \mathbb{E}(I_n^a(\Lambda)) = \mathbb{E}(I_n^b(\Lambda))$$

and

$$\mathbb{E}(|\mathrm{tr}(H_l^n)|^2) = \mathbb{E}(\mathrm{tr}(H_l^{-n})\mathrm{tr}(H_l^n)) = \mathbb{E}(I_n^a(\Lambda)I_n^b(\Lambda)).$$

We will show that, for all $n \in \mathbb{N}$, in the limit $N \rightarrow \infty$, uniformly in $a \in (0, T)$,

$$\mathbb{E}(I_n^a(\Lambda)) \rightarrow I_n^a, \quad \mathbb{E}(I_n^a(\Lambda)I_n^b(\Lambda)) \rightarrow I_n^a I_n^b. \quad (31)$$

Then

$$\mathbb{E}(\mathrm{tr}(H_l^n)) \rightarrow I_n^a, \quad \mathbb{E}(|\mathrm{tr}(H_l^n)|^2) \rightarrow |I_n^a|^2$$

so

$$\mathbb{E}(|\mathrm{tr}(H_l^n) - I_n^a|^2) = \mathbb{E}(|\mathrm{tr}(H_l^n)|^2) - 2\mathbb{E}(\mathrm{tr}(H_l^n))I_n^a + |I_n^a|^2 \rightarrow 0$$

as required.

The following estimates hold for $|w| \leq 1/2$

$$|\mathrm{Log}(1+w)| \leq 2|w|, \quad |\mathrm{Log}(1+w) - w| \leq |w|^2.$$

We apply these estimates with $w = n/(N(z-x))$, for $N \geq 2n$ and for points z on the contour $\gamma_{R \vee \Lambda^*}$ and x in the support of μ_Λ , to obtain

$$|G_\Lambda^{N/n}(z)| \leq 2, \quad |G_\Lambda^{N/n}(z) - G_\Lambda(z)| \leq n/N$$

where

$$G_\Lambda(z) = \int_{\mathbb{R}} \frac{\mu_\Lambda(dx)}{z-x}.$$

Note that γ_R has length $4R + 2\pi$. By some straightforward estimation, on Ω_R^c ,

$$|I_n^a(\Lambda)| \leq \frac{1}{2\pi n} (4\Lambda^* + 2\pi) e^{nT(\Lambda^*+1)+2n}$$

while, on Ω_R ,

$$|I_n^a(\Lambda)| \leq \frac{1}{2\pi n} (4R + 2\pi) e^{nT(R+1)+2n}.$$

Then, by the estimate (30), uniformly in $a \in (0, T)$,

$$\mathbb{E}(|I_n^a(\Lambda)|1_{\Omega_R^c}) \rightarrow 0, \quad \mathbb{E}(|I_n^a(\Lambda)I_n^b(\Lambda)|1_{\Omega_R^c}) \rightarrow 0.$$

On the other hand, on the event Ω_R , we have

$$I_n^a(\Lambda) = \frac{e^{-an^2/(2N)}}{2\pi in} \int_{\gamma_R} \exp\{-n(az - G_\Lambda^{N/n}(z))\} dz.$$

Hence, the weak limit (29) implies that, uniformly in $a \in (0, T)$,

$$I_n^a(\Lambda)1_{\Omega_R} \rightarrow \frac{1}{2\pi in} \int_{\gamma_R} \exp\{-n(az - G_T(z))\} dz = I_n^a$$

in probability, and so

$$\mathbb{E}(I_n^a(\Lambda)1_{\Omega_R}) \rightarrow I_n^a, \quad \mathbb{E}(I_n^a(\Lambda)I_n^b(\Lambda)1_{\Omega_R}) \rightarrow I_n^a I_n^b.$$

The desired limits (31) now follow. \square

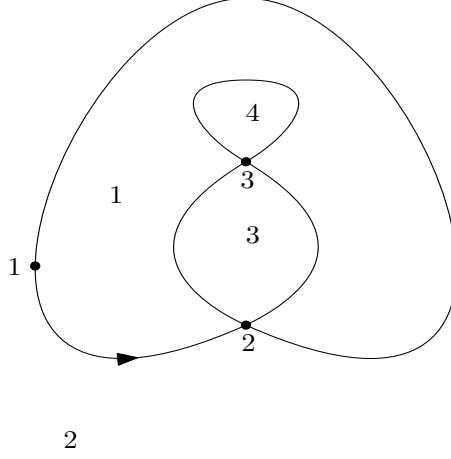


Fig. 4. A labelled embedded graph of a regular loop (see definition on page 8), with the standard labelling written on vertices and faces.

4. Makeenko–Migdal equations

Our aim in this section is to prove Proposition 2.6. For this, our main tool will be the Makeenko–Migdal equations. In order to formulate these precisely, we first give a description of the set of regular loops modulo area-preserving homeomorphisms of \mathbb{S}_T . This allows to reduce our analysis to a series of finite-dimensional simplices, each representing the possible vectors of face-areas for a given combinatorial graph. We show that the Makeenko–Migdal equations allow us to move area between faces of a regular loop provided only that the total area and the total winding number are conserved. This finally allows an inductive scheme to bootstrap the convergence we have shown for simple loops to all regular loops.

4.1. Combinatorial planar graphs and loops. Recall from Subsection 2.1 the notion of a labelled embedded graph. Given two labelled embedded graphs $\mathbb{G} = (e_1, \dots, e_m)$ and $\mathbb{G}' = (e'_1, \dots, e'_m)$, let us write $\mathbb{G} \sim \mathbb{G}'$ if there is an orientation-preserving homeomorphism θ of \mathbb{S}_T such that $e'_j = \theta \circ e_j$ for all j . Further, let us write $\mathbb{G} \approx \mathbb{G}'$ if θ may be chosen to be area-preserving. Then \sim and \approx are equivalence relations on the set of labelled embedded graphs. We will call the equivalence class of \mathbb{G} under \sim the *combinatorial graph* associated to \mathbb{G} .

We define a *standard labelling* of the vertices and faces of \mathbb{G} as follows. Consider the sequence of vertices $(\underline{e}_1, \bar{e}_1, \dots, \underline{e}_m, \bar{e}_m)$ and write $V = (v_1, \dots, v_q)$ for the subsequence obtained by dropping any vertex which has already appeared. Similarly consider the sequence of faces $(l(e_1), r(e_1), \dots, l(e_m), r(e_m))$, where $l(e_j)$ and $r(e_j)$ are the connected components of $\mathbb{S}_T \setminus \{e_1^*, \dots, e_m^*\}$ to the left and right of e_j . Then write $F = (f_1, \dots, f_p)$ for the subsequence obtained by dropping any face which has already appeared. See Figure 4 for an example. Set

$$\mathcal{V} = \{1, \dots, q\}, \quad \mathcal{E} = \{1, \dots, m\}, \quad \mathcal{F} = \{1, \dots, p\}.$$

The combinatorial graph associated to \mathbb{G} is then characterized⁵ by the integers q, m, p and the functions $s, t : \mathcal{E} \rightarrow \mathcal{V}$ and $l, r : \mathcal{E} \rightarrow \mathcal{F}$ given by

- (a) $s(j) = i$ if v_i is the starting point of e_j ,

⁵ To see this, given \mathbb{G}' with the same combinatorial data, we can first define homeomorphisms $e_j^* \rightarrow e_j'^*$ by parametrization at constant speed, then extend the resulting homeomorphisms of face-boundaries to homeomorphisms of closed faces to obtain a homeomorphism of \mathbb{S}_T .

- (b) $t(j) = i$ if v_i is the terminal point of e_j ,
- (c) $l(j) = k$ if f_k is the face to the left of e_j ,
- (d) $r(j) = k$ if f_k is the face to the right of e_j .

We call any quadruple $\mathcal{G} = (s, t, l, r)$ which arises in this way a *combinatorial planar graph*. We freely identify \mathcal{G} with the corresponding equivalence class of labelled embedded graphs.

Given a combinatorial planar graph \mathcal{G} , consider the simplex

$$\Delta_{\mathcal{G}}(T) = \{(a_1, \dots, a_p) : a_k > 0 \text{ for all } k \text{ and } a_1 + \dots + a_p = T\}.$$

Given a labelled embedded graph $\mathbb{G} \in \mathcal{G}$, define the *face-area vector* $a(\mathbb{G}) = (a_1, \dots, a_p)$ by

$$a_k = \text{area}(f_k).$$

Then $a(\mathbb{G}) \in \Delta_{\mathcal{G}}(T)$. For $a \in \Delta_{\mathcal{G}}(T)$, set

$$\mathcal{G}(a) = \{\mathbb{G} \in \mathcal{G} : a(\mathbb{G}) = a\}.$$

The sets $\mathcal{G}(a)$ are then the equivalence classes of the relation \approx . We call a sequence $\mathfrak{l}_0 = ((j_1, \varepsilon_1), \dots, (j_r, \varepsilon_r))$ in $\mathcal{E} \times \{-1, 1\}$ a *loop in \mathcal{G}* if

$$t(j_k, \varepsilon_k) = s(j_{k+1}, \varepsilon_{k+1}) \quad (32)$$

for $k = 1, \dots, r$, where $j_{r+1} = j_1$ and $\varepsilon_{r+1} = \varepsilon_1$ and where

$$s(j, \varepsilon) = t(j, -\varepsilon) = \begin{cases} s(j), & \text{if } \varepsilon = 1, \\ t(j), & \text{if } \varepsilon = -1. \end{cases}$$

The condition (32) means that, in any labelled embedded graph $\mathbb{G} = (e_1, \dots, e_m) \in \mathcal{G}$, we can concatenate the sequence of edges $(e_{j_1}^{\varepsilon_1}, e_{j_2}^{\varepsilon_2}, \dots, e_{j_r}^{\varepsilon_r})$ to form a loop

$$l_0 = e_{j_1}^{\varepsilon_1} e_{j_2}^{\varepsilon_2} \dots e_{j_r}^{\varepsilon_r}.$$

We call the loop l_0 so obtained the *drawing* of \mathfrak{l}_0 in \mathbb{G} . Note that the sequence

$$\mathfrak{l}^{-1} = ((j_r, -\varepsilon_r), \dots, (j_1, -\varepsilon_1))$$

is then also a loop in \mathcal{G} , whose drawing in \mathbb{G} is the reversal l^{-1} of l . Note also the obvious notion of concatenation for loops in \mathcal{G} .

In the case of interest to us, \mathcal{G} will be the combinatorial graph of the labelled embedded graph $\mathbb{G} = (e_1, \dots, e_m)$ of a regular loop l . We write then $a(l)$ for $a(\mathbb{G})$. If l has n self-intersections, we have $q = n + 1$, $m = 2n + 1$ and, by Euler's relation, $p = n + 2$. Note that the set of self-intersections is given in the standard labelling by $\{v_i : i \in \mathcal{I}\}$, where $\mathcal{I} = \{2, 3, \dots, n + 1\}$. We recover l as the drawing in \mathbb{G} of the loop

$$\mathfrak{l} = ((1, 1), \dots, (2n + 1, 1))$$

in \mathcal{G} . We call the pair $(\mathcal{G}, \mathfrak{l})$ a *combinatorial planar loop*. For each $n \geq 0$, there are only finitely many combinatorial loops with n self-intersections. We will write abusively \mathfrak{l} for $(\mathcal{G}, \mathfrak{l})$, $\Delta_{\mathfrak{l}}(T)$ for $\Delta_{\mathcal{G}}(T)$ and $\mathfrak{l}(a)$ for $\mathcal{G}(a)$. Given a loop \mathfrak{l}_0 in \mathcal{G} , it may be that the drawing l_0 of \mathfrak{l}_0 in \mathbb{G} is a regular loop. We could then consider the combinatorial loop associated to l_0 , without reference to its relation to \mathcal{G} . We will therefore need to make clear when such a combinatorial loop is to be considered in the context of a larger combinatorial graph. We shall also write $l_0 \in \mathfrak{l}(a)$ whenever $l_0 \in \text{Loop}(\mathbb{S}_T)$ is a drawing of \mathfrak{l} in a graph belonging to $\mathfrak{l}(a)$, for some area vector a . Given $n \geq 0$ and a function $\Phi : \text{Loop}_n(\mathbb{S}_T) \rightarrow \mathbb{C}$ which is invariant under area-preserving homeomorphisms, for any combinatorial planar loop \mathfrak{l} having n self-intersections, we can define a *quotient map* $\phi_{\mathfrak{l}} : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{C}$ by setting

$$\phi_{\mathfrak{l}}(a) = \Phi(l)$$

where l is any loop in $\mathfrak{l}(a)$. We say that Φ is *uniformly continuous in area* if the map $\phi_{\mathfrak{l}} : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{C}$ is uniformly continuous for all such combinatorial loops \mathfrak{l} .

4.2. Generalized Makeenko–Migdal equations. Let \mathfrak{l} be a combinatorial planar loop. Write m and p for the numbers of edges and faces in the associated combinatorial graph. Let $H = (H_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ be a Yang–Mills holonomy field in $U(N)$.

Proposition 4.1 *Let $f : U(N)^m \rightarrow \mathbb{C}$ be a continuous bounded function. Then we can define a function $E(f) : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{C}$ by*

$$E(f)(a) = \frac{1}{p_T(1)} \int_{U(N)^m} f(g_1, \dots, g_m) \prod_{k=1}^p p_{a_k}(\tilde{g}_k) \prod_{j=1}^m dg_j$$

where \tilde{g}_k is the holonomy around f_k obtained from the edge holonomies g_1, \dots, g_m . Moreover $E(f)$ is uniformly continuous on $\Delta_{\mathfrak{l}}(T)$ and

$$E(f)(a) = \mathbb{E}(f(H_{e_1}, \dots, H_{e_m}))$$

for $a \in \Delta_{\mathfrak{l}}(T)$, whenever $\mathbb{G} = (e_1, \dots, e_m)$ is a labelled embedded graph with $\mathbb{G} \in \mathfrak{l}(a)$.

Proof. The function $E(f)$ is well defined because $p_a(g) = p_a(g^{-1}) = p_a(hgh^{-1})$, which ensures that the right-hand side does not depend on the choices of starting point and direction for the loop holonomy \tilde{g}_k . It will suffice to show uniform continuity on each of the sets

$$\Delta_k = \{a \in \Delta_{\mathfrak{l}}(T) : a_k \geq T/p\}.$$

Then, by symmetry, it will suffice to consider the case $k = p$. Given a family of edge holonomies $g = (g_1, \dots, g_m) \in U(N)^m$, for each face f_k , choose an adjacent edge and denote by \tilde{g}_k the holonomy around f_k starting from that edge. Fix also edge labels i_1, \dots, i_{m-p+1} such that the edges $e_{i_1}, \dots, e_{i_{m-p+1}}$ form a spanning tree of the associated graph. Define, for $k = 1, \dots, p-1$ and $j = 1, \dots, m-p+1$,

$$b_k = \tilde{g}_k, \quad h_j = g_{i_j}.$$

Then the map $g \mapsto (b, h) : U(N)^m \rightarrow U(N)^m$ preserves the m -fold product of Haar measure. Moreover, the edge holonomies g_1, \dots, g_m and the loop holonomy \tilde{g}_p for f_p are given by finite products of b_1, \dots, b_{p-1} and h_1, \dots, h_{m-p+1} and their inverses. See [41, Proposition 2.4.2]. Hence we have

$$\begin{aligned} E(f)(a) &= \frac{1}{p_T(1)} \int_{U(N)^m} f(g_1(b, h), \dots, g_m(b, h)) p_{a_p}(\tilde{g}_p(b, h)) \prod_{k=1}^{p-1} p_{a_k}(b_k) db_k \prod_{j=1}^{m-p+1} dh_j \\ &= \frac{1}{p_T(1)} \mathbb{E} \int_{U(N)^{m-p+1}} f(g_1(B, h), \dots, g_m(B, h)) p_{a_p}(\tilde{g}_p(B, h)) \prod_{j=1}^{m-p+1} dh_j \end{aligned} \quad (33)$$

where $B = (B_{a_1}^1, \dots, B_{a_{p-1}}^{p-1})$ and B^1, \dots, B^{p-1} are independent Brownian motions in G starting from 1. Now, on Δ_p , we have $a_p \geq T/p$, so the claimed uniform continuity follows from standard continuity estimates for Brownian motion and the heat kernel in G . \square

For $i \in \{1, \dots, m\}$ and $g \in U(N)$, define maps $R_{i,g}$ and $\hat{R}_{i,g}$ on $U(N)^m$ by

$$\begin{aligned} R_{i,g}(h_1, \dots, h_m) &= (h_1, \dots, h_i g, \dots, h_m), \\ \hat{R}_{i,g}(h_1, \dots, h_m) &= (h_1, \dots, g^{-1} h_i, \dots, h_m). \end{aligned}$$

For $i \in \{1, \dots, m\}$ and $X \in \mathfrak{u}(N)$, define a differential operator \mathcal{L}_X^i on $U(N)^m$ by

$$\mathcal{L}_X^i(f) = \left. \frac{d}{dt} \right|_{t=0} f \circ R_{i, e^{tX}}.$$

Choose an orthonormal basis $(X_n : n = 1, \dots, N^2)$ for $\mathfrak{u}(N)$ (with inner product (1)) and, for $i, j \in \{1, \dots, m\}$, define

$$\Delta_{i,j}(f) = \sum_n \mathcal{L}_{X_n}^i \circ \mathcal{L}_{X_n}^j(f).$$

The operator $\Delta_{i,j}$ does not depend on the choice of orthonormal basis.

A function $f : U(N)^m \rightarrow \mathbb{C}$ is said to have *extended gauge invariance* if, for all $g \in U(N)$ and for $i = 1, \dots, m-1$,

$$f \circ \hat{R}_{i,g} \circ R_{i+1,g} = f.$$

Thus we require

$$f(h_1, \dots, g^{-1}h_i, h_{i+1}g, \dots, h_m) = f(h_1, \dots, h_i, h_{i+1}, \dots, h_m).$$

Recall that

$$\Delta_{\mathfrak{l}}(T) = \{(a_1, \dots, a_{n+2}) : a_k > 0 \text{ for all } k \text{ and } a_1 + \dots + a_{n+2} = T\}.$$

Write \mathcal{I} for the set of intersection labels and \mathcal{F} for the set of face labels in the combinatorial graph \mathcal{G} of \mathfrak{l} , as usual. For $i \in \mathcal{I}$, define a (constant) vector field Ξ_i on $\Delta_{\mathfrak{l}}(T)$ as follows. Choose $\mathbb{G} \in \mathcal{G}$ and write l for the drawing of \mathfrak{l} in \mathbb{G} . In the standard labelling of \mathbb{G} , the vertex v_i is a self-intersection of l . Write (k_1, k_2, k_3, k_4) for the labels of the faces found on making a small anti-clockwise circuit around v_i , starting in the face f_{k_1} adjacent to two outgoing edges, and in the corner adjacent to those edges. Note that the case $k_1 = k_3$ can arise, but the condition that we start in the corner adjacent to the outgoing edges allows us to specify the sequence (k_1, k_2, k_3, k_4) uniquely in any case. In the example of Figure 4, if $i = 2$, then $k_1 = k_3 = 1$, $k_2 = 3$ and $k_4 = 2$. This sequence does not depend on the choice of \mathbb{G} . Set

$$\Xi_i = \partial_{k_1} - \partial_{k_2} + \partial_{k_3} - \partial_{k_4} \quad (34)$$

where $\partial_k = \partial/\partial a_k$ denotes the elementary vector field in direction k . Note that Ξ_i is tangent to the simplex $\Delta_{\mathfrak{l}}(T)$.

The following theorem is a specialization of a result of Driver, Gabriel, Hall and Kemp [18, Theorem 2], which generalizes a formulation of Lévy [42].

Theorem 4.2 *Let $f : U(N)^m \rightarrow \mathbb{C}$ be a C^∞ function having extended gauge invariance. Then, for all $i \in \mathcal{I}$, the function $E(f)$ has a directional derivative on $\Delta_{\mathfrak{l}}(T)$ in direction Ξ_i given by*

$$\Xi_i E(f) = -E(\Delta_{j_1, j_2}(f))$$

where j_1, j_2 are determined by $s(j_1) = s(j_2) = i$.

4.3. Makeenko–Migdal equations for Wilson loops. Given a loop $\mathfrak{l}_0 = ((j_1, \varepsilon_1), \dots, (j_r, \varepsilon_r))$ in \mathcal{G} , we can define a continuous bounded function $W_{\mathfrak{l}_0} : U(N)^m \rightarrow \mathbb{C}$ by

$$W_{\mathfrak{l}_0}(h_1, \dots, h_m) = \text{tr}(h_{j_r}^{\varepsilon_r} \dots h_{j_1}^{\varepsilon_1}).$$

Given a sequence of loops $(\mathfrak{l}_1, \dots, \mathfrak{l}_k)$ in \mathcal{G} , define the *Wilson loop function*

$$\phi_{\mathfrak{l}_1, \dots, \mathfrak{l}_k}^N : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{C}$$

by

$$\phi_{\mathfrak{l}_1, \dots, \mathfrak{l}_k}^N = E(W_{\mathfrak{l}_1} \dots W_{\mathfrak{l}_k}).$$

Then $\phi_{\mathfrak{l}_1, \dots, \mathfrak{l}_k}^N$ is uniformly continuous and, for all $a \in \Delta_{\mathfrak{l}}(T)$ and all $\mathbb{G} \in \mathfrak{l}(a)$,

$$\phi_{\mathfrak{l}_1, \dots, \mathfrak{l}_k}^N(a) = \mathbb{E}(\text{tr}(H_{\mathfrak{l}_1}) \dots \text{tr}(H_{\mathfrak{l}_k})) \quad (35)$$

where l_1, \dots, l_k are the drawings of l_1, \dots, l_k in \mathbb{G} .

For $i \in \mathcal{I}$, we obtain two regular loops l_i and \hat{l}_i by splitting l at v_i , that is, by following the two outgoing strands of l from v_i until their first return to v_i . In one case we will pass through the endpoint of l and begin another circuit of l until we reach v_i . Write l_i and \hat{l}_i for the loops in \mathcal{G} whose drawings in \mathbb{G} are l_i and \hat{l}_i , which do not depend on the choice of \mathbb{G} . Then set

$$[l]_i = l_i \hat{l}_i l_i^{-1} \hat{l}_i^{-1}, \quad [\hat{l}]_i = \hat{l}_i l_i \hat{l}_i^{-1} l_i^{-1}$$

where l_i^{-1}, \hat{l}_i^{-1} denote the reversals of l_i, \hat{l}_i and the right-hand sides are understood as concatenations.

Proposition 4.3 (Makeenko–Migdal equations for Wilson loops) *The functions ϕ_l^N and $\phi_{l, l^{-1}}^N$ have directional derivatives in $\Delta_l(T)$ in direction Ξ_i given by*

$$\Xi_i \phi_l^N = \phi_{l_i, \hat{l}_i}^N, \quad \Xi_i \phi_{l, l^{-1}}^N = \phi_{l_i, \hat{l}_i, l^{-1}}^N + \phi_{l_i, l_i^{-1}, \hat{l}_i^{-1}}^N - N^{-2}(\phi_{[l]_i}^N + \phi_{[\hat{l}]_i}^N).$$

Proof. We give details only for $\phi_{l, l^{-1}}^N$. The simpler argument for ϕ_l^N will then be obvious. The argument for ϕ_l^N already appeared after Theorem 2.6 in [19] and in Section 9.2 of [42]. Given $\mathbb{G} = (e_1, \dots, e_m) \in \mathcal{G}$, set $l = e_1 \dots e_m$, so l is the drawing of l in \mathbb{G} . Given $h = (h_1, \dots, h_m) \in U(N)^m$, there is a unique multiplicative function

$$(h_\gamma : \gamma \in \text{Path}(\mathbb{G})) \in \text{Mult}(\text{Path}(\mathbb{G}), U(N))$$

such that $h_{e_j} = h_j$ for all j . Then $\phi_{l, l^{-1}}^N = E(f)$, where $f = |W_l|^2$ and

$$W_l(h_1, \dots, h_m) = \text{tr}(h_l) = \text{tr}(h_m \dots h_1).$$

Note that W_l has extended gauge invariance and so also does f . We can write $l_i = e\gamma$ and $\hat{l}_i = \hat{e}\hat{\gamma}$, where $e = e_{j_1}, \hat{e} = e_{j_2}, s(j_1) = s(j_2) = i$ and $\gamma, \hat{\gamma} \in \text{Path}(\mathbb{G})$. Then

$$f(h) = \text{tr}(h_l) \text{tr}(h_l^{-1}) = \text{tr}(h_{\hat{\gamma}} h_{\hat{e}} h_{\gamma} h_e) \text{tr}(h_e^{-1} h_{\gamma}^{-1} h_{\hat{e}}^{-1} h_{\hat{\gamma}}^{-1}).$$

For $X \in \mathfrak{u}(N)$,

$$\begin{aligned} \mathcal{L}_X^{j_1} \circ \mathcal{L}_X^{j_2}(f)(h) &= \text{tr}(h_{\hat{\gamma}} h_{\hat{e}} X h_{\gamma} h_e X) \text{tr}(h_l^{-1}) + \text{tr}(h_l) \text{tr}(X h_e^{-1} h_{\gamma}^{-1} X h_{\hat{e}}^{-1} h_{\hat{\gamma}}^{-1}) \\ &\quad - \text{tr}(h_{\hat{\gamma}} h_{\hat{e}} h_{\gamma} h_e X) \text{tr}(h_e^{-1} h_{\gamma}^{-1} X h_{\hat{e}}^{-1} h_{\hat{\gamma}}^{-1}) \\ &\quad - \text{tr}(h_{\hat{\gamma}} h_{\hat{e}} X h_{\gamma} h_e) \text{tr}(X h_e^{-1} h_{\gamma}^{-1} h_{\hat{e}}^{-1} h_{\hat{\gamma}}^{-1}). \end{aligned}$$

Write $E_{j,k}$ for the elementary matrix with a 1 in the (j, k) -entry. Set

$$X_{j,j} = iE_{j,j}/\sqrt{N}, \quad X_{j,k} = \begin{cases} (E_{j,k} - E_{k,j})/\sqrt{2N}, & \text{for } j < k, \\ i(E_{j,k} + E_{k,j})/\sqrt{2N}, & \text{for } j > k. \end{cases}$$

Then $\{X_{j,k} : j, k = 1, \dots, N\}$ is an orthonormal basis in $\mathfrak{u}(N)$. A simple calculation gives the standard identity

$$\sum_{j,k=1}^N X_{j,k} \otimes X_{j,k} = -\frac{1}{N} \sum_{j,k=1}^N E_{j,k} \otimes E_{k,j}.$$

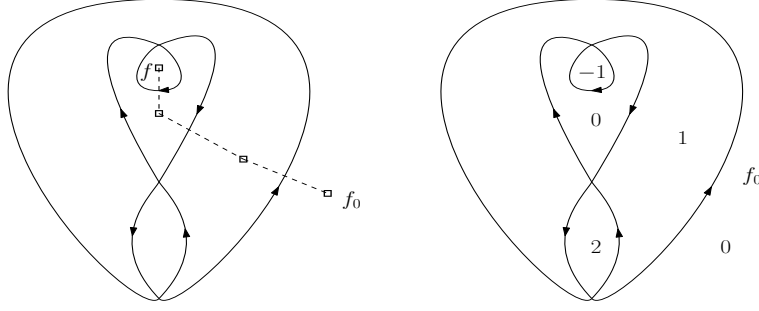


Fig. 5. A track between two faces f_0 and f is drawn with dashed lines. The value of the winding number for the choice of f_0 is printed on each face.

We sum to obtain

$$\begin{aligned}
 -\Delta_{j_1, j_2}(f)(h) &= \text{tr}(h_{\hat{\gamma}} h_{\hat{e}}) \text{tr}(h_{\gamma} h_e) \text{tr}(h_l^{-1}) + \text{tr}(h_l) \text{tr}(h_e^{-1} h_{\gamma}^{-1}) \text{tr}(h_{\hat{e}}^{-1} h_{\hat{\gamma}}^{-1}) \\
 &\quad - \frac{1}{N^2} \text{tr}(h_{\hat{\gamma}} h_{\hat{e}} h_{\gamma} h_e h_{\hat{e}}^{-1} h_{\gamma}^{-1} h_e^{-1} h_{\hat{\gamma}}^{-1}) \\
 &\quad - \frac{1}{N^2} \text{tr}(h_{\gamma} h_e h_{\hat{\gamma}} h_{\hat{e}} h_e^{-1} h_{\gamma}^{-1} h_{\hat{e}}^{-1} h_{\hat{\gamma}}^{-1})
 \end{aligned}$$

and hence, by Theorem 4.2,

$$\Xi_i \phi_{l, l^{-1}}^N = -E(\Delta_{j_1, j_2}(f)) = \phi_{l_i, \hat{l}_i, l^{-1}}^N + \phi_{l_i, l_i^{-1}, \hat{l}_i^{-1}}^N - N^{-2}(\phi_{[\hat{l}]_i}^N + \phi_{[\hat{l}]_i}^N).$$

□

4.4. Makeenko–Migdal vectors and the winding number. Let $l \in \text{Loop}(\mathbb{S}_T)$ be a regular loop and let $\mathbb{G} = (V, E, F)$ be the associated labelled embedded graph. For any pair of faces $f_0, f_* \in F$ and points $x_0 \in f_0, x \in f_*$, the set $\mathbb{S}_T \setminus \{x_0, x\}$ can be retracted to a simple closed curve s in \mathbb{S}_T which winds positively around x_* and negatively around x_0 . Furthermore, there is a unique $n_l(f_0, f_*) \in \mathbb{Z}$ such that l is homotopic within $\mathbb{S}_T \setminus \{x_0, x\}$ to $s^{n_l(f_0, f_*)}$. The integer $n_l(f_0, f_*)$ does not depend on the choice of x_0, x_* but only on f_0, f_* . Setting $n_l(f, f) = 0$ for any $f \in F$, this defines a skew symmetric function $n_l : F^2 \rightarrow \mathbb{Z}$. Fixing an orientation preserving homeomorphism from $\mathbb{S}_T \setminus \{x_0\}$ to \mathbb{R}^2 , for each face $f \in F$, $n_l(f_0, f)$ is the winding number of the image of l around the image of f in the plane \mathbb{R}^2 . This number can be computed as follows. Given a *track* from f_0 to f , comprising edges e_1, \dots, e_k and faces f_1, \dots, f_k such that $f_k = f$ and e_j is adjacent to both f_{j-1} and f_j for all j , we have

$$n_l(f_0, f) = L(f) - R(f)$$

where $L(f)$ and $R(f)$ are the numbers of edges e_j with f_j on the left and right respectively. (The notation here does not refer to the standard labelling of \mathbb{G} .) This construction yields the following observation: for any $f_0, f_* \in F$,

$$n_l(f_0, f) + n_l(f, f_*) = n_l(f_0, f_*) = -n_l(f_*, f_0).$$

It follows that the function $f \mapsto n_l(f_0, f)$ depends on the choice of face f_0 only through the addition of a constant. We shall denote it abusively as well by n_l and call it the *winding number function* of l . See Figure 5 for an example.

The winding number is invariant under orientation-preserving homeomorphisms of \mathbb{S}_T , so, using the notations introduced in section 4.1, we obtain also a function

$$n_{\mathfrak{l}} : \mathcal{F} \rightarrow \mathbb{Z}$$

determined by the associated combinatorial loop \mathfrak{l} , also defined up to an additive constant, by setting

$$n_{\mathfrak{l}}(k) = n_{\mathfrak{l}}(f)$$

where f is the k th face in the standard labelling of \mathbb{G} .

The following lemma is a reformulation of a lemma of Lévy [42, Lemma 6.28]. See also Dahlqvist [13, Lemma 21]. We give a slightly different proof, relying on properties of the winding number in place of a dimension-counting argument. The prior results were stated for the whole plane, while ours applied to the sphere, but this make little difference to the argument.

Lemma 4.4 *There is an orthogonal direct sum decomposition*

$$\mathbb{R}^{\mathcal{F}} = \mathfrak{m}_{\mathfrak{l}} \oplus \mathfrak{n}_{\mathfrak{l}}$$

where

$$\mathfrak{m}_{\mathfrak{l}} = \text{span}\{\Xi_i : i \in \mathcal{I}\}, \quad \mathfrak{n}_{\mathfrak{l}} = \text{span}\{1, n_{\mathfrak{l}}\}.$$

Proof. Note first that $1^T \Xi_i = 1 - 1 + 1 - 1 = 0$ for all i . Let $i \in \mathcal{I}$. Write k_1, k_2, k_3, k_4 for the faces at i , listed anticlockwise starting from the face k_1 adjacent to both outgoing edges. Then the values of $n_{\mathfrak{l}}$ at k_1, k_2, k_3, k_4 are given respectively by $n, n+1, n, n-1$ for some n , so

$$n_{\mathfrak{l}}^T \Xi_i = n_{\mathfrak{l}}(k_1) - n_{\mathfrak{l}}(k_2) + n_{\mathfrak{l}}(k_3) - n_{\mathfrak{l}}(k_4) = 0.$$

Hence, if $\alpha \in \mathfrak{m}_{\mathfrak{l}}$, then $1^T \alpha = 0$ and $n_{\mathfrak{l}}^T \alpha = 0$.

Suppose on the other hand that $\alpha \in \mathfrak{n}_{\mathfrak{l}}^\perp$. Consider the 1-forms (of the dual graph) $d\alpha$ and $dn_{\mathfrak{l}}$, given by

$$d\alpha(j) = \alpha(l(j)) - \alpha(r(j)), \quad dn_{\mathfrak{l}}(j) = n_{\mathfrak{l}}(l(j)) - n_{\mathfrak{l}}(r(j)), \quad j \in \mathcal{E}.$$

Then $dn_{\mathfrak{l}}(j) = 1$ for all j . On the other hand, for $j = 1, \dots, m-1$, there is an $i_j \in \mathcal{I}$ such that $t(j) = i_j = s(j+1)$, so

$$d\alpha(j) - d\alpha(j+1) = \pm \Xi_{i_j}^T \alpha = 0.$$

Hence $d\alpha = c_1 dn_{\mathfrak{l}}$ and so $\alpha = c_1 n_{\mathfrak{l}} + c_2$ for some constants c_1, c_2 . \square

Note that $\Delta_{\mathfrak{l}}(T)$ is convex, and that, by counting dimensions, the vectors $\{\Xi_i : i \in \mathcal{I}\}$ are linearly independent. We deduce from these facts, and the preceding lemma the following proposition. Write $\overline{\Delta_{\mathfrak{l}}(T)}$ for the closure of $\Delta_{\mathfrak{l}}(T)$ in \mathbb{R}^p .

Proposition 4.5 *Let $a \in \Delta_{\mathfrak{l}}(T)$ and $a' \in \overline{\Delta_{\mathfrak{l}}(T)}$. Set $w = a' - a$. Then $a + tw \in \Delta_{\mathfrak{l}}(T)$ for all $t \in [0, 1)$. Moreover, there exists $\alpha \in \mathbb{R}^{\mathcal{I}}$ such that*

$$w = \sum_{i \in \mathcal{I}} \alpha_i \Xi_i$$

if and only if

$$\sum_{k \in \mathcal{F}} a_k n_{\mathfrak{l}}(k) = \sum_{k \in \mathcal{F}} a'_k n_{\mathfrak{l}}(k).$$

Moreover, in this case, α is uniquely determined by w and

$$\sum_{i \in \mathcal{I}} |\alpha_i| \leq C_{\mathfrak{l}} \sum_{k \in \mathcal{F}} |a_k - a'_k|$$

for some constant $C_{\mathfrak{l}} < \infty$ depending only on \mathfrak{l} .

4.5. *Proof of Proposition 2.6.* We will show inductively that the following statements hold for all $n \geq 0$. Firstly, for all combinatorial planar loops \mathfrak{l} with no more than n self-intersections, there is a continuous function

$$\phi_{\mathfrak{l}} : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{R}$$

such that, uniformly on $\Delta_{\mathfrak{l}}(T)$ as $N \rightarrow \infty$,

$$\phi_{\mathfrak{l}}^N \rightarrow \phi_{\mathfrak{l}}, \quad \phi_{\mathfrak{l}, \mathfrak{l}-1}^N \rightarrow (\phi_{\mathfrak{l}})^2.$$

Secondly, the restriction of the master field Φ_T to $\text{Loop}_n(\mathbb{S}_T)$ is the unique function $\text{Loop}_n(\mathbb{S}_T) \rightarrow \mathbb{C}$ with the following properties: it is invariant under area-preserving diffeomorphisms, for any combinatorial planar loop \mathfrak{l} having at most n self-intersections, the quotient map $\phi_{\mathfrak{l}} : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{R}$ is uniformly continuous and Φ_T satisfies the Makeenko–Migdal equations (10) and the estimate (13).

For $a \in \Delta_{\mathfrak{l}}(T)$ and $l \in \mathfrak{l}(a)$,

$$\mathbb{E}(|\text{tr}(H_l) - \phi_{\mathfrak{l}}(a)|^2) = \phi_{\mathfrak{l}, \mathfrak{l}-1}^N(a) - \phi_{\mathfrak{l}}^N(a)^2 + (\phi_{\mathfrak{l}}^N(a) - \phi_{\mathfrak{l}}(a))^2$$

so the first statement implies that, as $N \rightarrow \infty$,

$$\text{tr}(H_l) \rightarrow \phi_{\mathfrak{l}}(a) = \Phi_T(l)$$

in L^2 , uniformly in $l \in \text{Loop}_n(\mathbb{S}_T)$. So the two statements suffice to prove Proposition 2.6.

For the simple combinatorial loop \mathfrak{s} , set

$$\phi_{\mathfrak{s}}(a, b) = \phi_T(1, a, b)$$

then $\phi_{\mathfrak{s}}$ is continuous on $\Delta_{\mathfrak{s}}(T)$ and, by Proposition 2.5, $\phi_{\mathfrak{s}}^N \rightarrow \phi_{\mathfrak{s}}$ and $\phi_{\mathfrak{s}, \mathfrak{s}-1}^N \rightarrow (\phi_{\mathfrak{s}})^2$ uniformly on $\Delta_{\mathfrak{s}}(T)$. There are no self-intersections, so no Makeenko–Migdal equations. For $(a, b) \in \Delta_{\mathfrak{s}}(T)$ and $s \in \mathfrak{s}(a, b)$,

$$\Phi_T(s) = \phi_{\mathfrak{s}}(a, b) = \phi_T(1, a, b).$$

Hence the desired statements hold for $n = 0$.

Let $n \geq 1$ and suppose inductively that the desired statements hold for $n - 1$. Let \mathfrak{l} be a combinatorial planar loop with n self-intersections. Choose faces k_0 and k_* of minimal and maximal winding number and set

$$n_* = n_{\mathfrak{l}}(k_*) - n_{\mathfrak{l}}(k_0).$$

Let $a = (a_1, \dots, a_{n+2}) \in \Delta_{\mathfrak{l}}(T)$. Recall that $a_0, a_* \in [0, T]$ are determined by

$$a_0 + a_* = T, \quad a_0 n_{\mathfrak{l}}(k_0) + a_* n_{\mathfrak{l}}(k_*) = \sum_{k=1}^{n+2} a_k n_{\mathfrak{l}}(k). \quad (36)$$

Then, by Proposition 4.5, there exists a unique $\alpha \in \mathbb{R}^{\mathcal{I}}$, with

$$\sum_{i \in \mathcal{I}} |\alpha_i| \leq 2C_{\mathfrak{l}}(T - a_{k_0} - a_{k_*}) \quad (37)$$

such that, for

$$a(t) = a + t \sum_{i \in \mathcal{I}} \alpha_i \Xi_i \quad (38)$$

we have $a(t) \in \Delta_{\mathfrak{l}}(T)$ for all $t \in [0, 1]$ and

$$a_{k_0}(1) = a_0, \quad a_{k_*}(1) = a_*.$$

By Proposition 4.3, the maps

$$t \mapsto \phi_l^N(a(t)), \quad t \mapsto \phi_{l,l^{-1}}^N(a(t))$$

are differentiable on $[0, 1)$, with

$$\frac{d}{dt} \phi_l^N(a(t)) = \sum_{i \in \mathcal{I}} \alpha_i \phi_{l_i, \hat{l}_i}^N(a(t))$$

and

$$\frac{d}{dt} \phi_{l,l^{-1}}^N(a(t)) = \sum_{i \in \mathcal{I}} \alpha_i \left(\phi_{l_i, \hat{l}_i, l^{-1}}^N + \phi_{l_i, l_i^{-1}, \hat{l}_i}^N - N^{-2}(\phi_{[l]_i}^N + \phi_{[\hat{l}]_i}^N) \right) (a(t)).$$

Here we have used the fact that the directional derivatives given by Proposition 4.3 are continuous on $\Delta_l(T)$ to guarantee differentiability in any linear combination of those directions. We integrate to obtain, for all $t \in [0, 1)$,

$$\phi_l^N(a(t)) = \phi_l^N(a) + \sum_{i \in \mathcal{I}} \int_0^t \alpha_i \phi_{l_i, \hat{l}_i}^N(a(s)) ds \quad (39)$$

and

$$\begin{aligned} & \phi_{l,l^{-1}}^N(a(t)) \\ &= \phi_{l,l^{-1}}^N(a) + \sum_{i \in \mathcal{I}} \alpha_i \int_0^t \left(\phi_{l_i, \hat{l}_i, l^{-1}}^N + \phi_{l_i, l_i^{-1}, \hat{l}_i}^N - N^{-2}(\phi_{[l]_i}^N + \phi_{[\hat{l}]_i}^N) \right) (a(s)) ds. \end{aligned} \quad (40)$$

Since ϕ_l^N and $\phi_{l,l^{-1}}^N$ extend continuously to $\overline{\Delta_l(T)}$ and the integrands on the right are bounded, these equations hold also for $t = 1$.

We shall now prove the following key identities

$$\phi_l^N(a(1)) = \phi_{s^{n_*}}^N(a_0, a_*), \quad \phi_{l,l^{-1}}^N(a(1)) = \phi_{s^{n_*}, s^{-n_*}}^N(a_0, a_*). \quad (41)$$

Choose a loop $l \in \mathcal{I}(a)$ such that the faces f_{k_0} and f_{k_*} of the associated embedded graph have a C^1 boundary. We will use a deformation $(l_t)_{t \in [0, 1]}$ of l constructed using a diffeomorphism from $\mathbb{S}_T \setminus (f_{k_0} \cup f_{k_*})$ to a cylinder and then contracting the cylinder to a circle. Since k_* and k_0 have maximal and minimal winding number, the pair (a_0, a_*) defined by (36) satisfies $a_0 \geq a_{k_0}$ and $a_* \geq a_{k_*}$. Hence, there is an area-preserving C^1 diffeomorphism

$$F : \mathbb{S}_T \setminus (f_{k_0} \cup f_{k_*}) \rightarrow (\mathbb{R}/\mathbb{Z}) \times [-a_0 + a_{k_0}, a_* - a_{k_*}]$$

where the right-hand side is endowed with Lebesgue measure. By re-basing the loop l if necessary, we may assume that the starting point $l(0)$ is not adjacent to f_0 or f_* . Write $F(l(\tau)) = (\theta(\tau), y(\tau))$. We can and do choose l and F so that $F(l(0)) = (0, 0)$ and so that $\dot{\theta}(\tau)$ makes only finitely many changes of sign. For $t \in [0, 1]$ and $(\theta, y) \in (\mathbb{R}/\mathbb{Z}) \times [-a_0 + a_{k_0}, a_* - a_{k_*}]$, define

$$C_t(\theta, y) = (\theta, (1-t)y)$$

and define a family $(l_t : t \in [0, 1])$ in $\text{Loop}(\mathbb{S}_T)$ by

$$l_t(\tau) = F^{-1} \circ C_t \circ F(l(\tau)). \quad (42)$$

Then $l_0 = l$ and $(l_t : t \in [0, 1])$ is continuous in length with fixed endpoints. Define

$$s(\tau) = F^{-1}(\tau, 0).$$

Then $s \in \text{Loop}_0(\mathbb{S}_T)$ and, since $F \circ l_1(\tau) = (\theta(\tau), 0)$, we have $l_1 \sim s^{n^*}$, where $n^* = \theta(1) \in \mathbb{Z}$ is the winding number of l . Then, by continuity in probability and invariance under reduction of the holonomy field,

$$\begin{aligned}\phi_l^N(a(1)) &= \lim_{t \uparrow 1} \phi_l^N(a(t)) = \lim_{t \uparrow 1} \mathbb{E}(\text{tr}(H_{l_t})) \\ &= \mathbb{E}(\text{tr}(H_{l_1})) = \mathbb{E}(\text{tr}(H_{s^{n^*}})) = \phi_{s^{n^*}}^N(a_0, a_*).\end{aligned}$$

We see similarly that $\phi_{l, l^{-1}}^N(a(1)) = \phi_{s^{n^*}, s^{-n^*}}^N(a_0, a_*)$.

Now, given (41), by Proposition 2.5,

$$\phi_{s^{n^*}}^N(a_1, a_2) \rightarrow \phi_T(n_*, a_1, a_2)$$

uniformly in $(a_1, a_2) \in \Delta_s(T)$. Write l_i and \hat{l}_i for the drawings of l_i and \hat{l}_i in \mathbb{G} for some $\mathbb{G} \in \mathcal{I}(a)$. Then

$$\phi_{l_i, \hat{l}_i}^N(a) = \mathbb{E}(\text{tr}(H_{l_i})\text{tr}(H_{\hat{l}_i})).$$

Since both l_i and \hat{l}_i have no more than $n - 1$ self-intersections, by the inductive hypothesis,

$$\text{tr}(H_{l_i}) \rightarrow \phi_{l_i}(a), \quad \text{tr}(H_{\hat{l}_i}) \rightarrow \phi_{\hat{l}_i}(a) \quad (43)$$

in L^2 , uniformly in $a \in \Delta_l(T)$. Hence

$$\phi_{l_i, \hat{l}_i}^N \rightarrow \phi_{l_i} \phi_{\hat{l}_i}$$

uniformly on $\Delta_l(T)$. Here we used the obvious submersions $\Delta_l(T) \rightarrow \Delta_{l_i}(T)$ and $\Delta_l(T) \rightarrow \Delta_{\hat{l}_i}(T)$ in evaluating ϕ_{l_i} and $\phi_{\hat{l}_i}$ on $\Delta_l(T)$. We let $N \rightarrow \infty$ in (39), first in the case $t = 1$ and then for $t \in (0, 1)$ to see that ϕ_l^N converges uniformly on $\overline{\Delta_l(T)}$ with continuous limit, ϕ_l say, satisfying, for all $t \in [0, 1]$,

$$\phi_l(a(t)) = \phi_l(a) + \sum_{i \in \mathcal{I}} \int_0^t \alpha_i \phi_{l_i}(a(s)) \phi_{\hat{l}_i}(a(s)) ds. \quad (44)$$

Using again (41) and Proposition 2.5, for $s \in s(a_1, a_2)$,

$$\phi_{s^n, s^{-n}}^N(a_1, a_2) = \mathbb{E}(|\text{tr}(H_s^n)|^2) \rightarrow \phi_T(n, a_1, a_2)^2$$

uniformly in $(a_1, a_2) \in \Delta_s(T)$ as $N \rightarrow \infty$. We have

$$\phi_{l_i, \hat{l}_i, l^{-1}}^N(a) = \phi_{l_i, l_i^{-1}, \hat{l}_i^{-1}}^N(a) = \mathbb{E}(\text{tr}(H_{l_i})\text{tr}(H_{\hat{l}_i})\text{tr}(H_{l^{-1}}))$$

and we have just shown that

$$\mathbb{E}(\text{tr}(H_{l^{-1}})) = \mathbb{E}(\text{tr}(H_l)) \rightarrow \phi_l(a)$$

uniformly in $a \in \Delta_l(T)$. In combination with (43), we deduce that, uniformly on $\Delta_l(T)$,

$$\phi_{l_i, \hat{l}_i, l^{-1}}^N \rightarrow \phi_{l_i} \phi_{\hat{l}_i} \phi_l.$$

Hence, on letting $N \rightarrow \infty$ in (40), first in the case $t = 1$ and then for $t \in (0, 1)$, we see that $\phi_{l, l^{-1}}^N$ converges uniformly on $\overline{\Delta_l(T)}$ with continuous limit, $\phi_{l, l^{-1}}$ say, satisfying, for all $t \in [0, 1]$,

$$\phi_{l, l^{-1}}(a(t)) = \phi_{l, l^{-1}}(a) + 2 \sum_{i \in \mathcal{I}} \int_0^t \alpha_i \phi_{l_i}(a(s)) \phi_{\hat{l}_i}(a(s)) \phi_l(a(s)) ds. \quad (45)$$

By differentiating (44) and (45), we see that

$$\frac{d}{dt} (\phi_{\mathfrak{l}, \mathfrak{l}^{-1}}(a(t)) - \phi_{\mathfrak{l}}(a(t))^2) = 0$$

so

$$\phi_{\mathfrak{l}, \mathfrak{l}^{-1}}(a) - \phi_{\mathfrak{l}}(a)^2 = \phi_{\mathfrak{l}, \mathfrak{l}^{-1}}(a(1)) - \phi_{\mathfrak{l}}(a(1))^2 = 0.$$

Thus the first of the desired statements holds for n .

We turn to the second statement. First we will show the claimed properties of the master field Φ_T on $\text{Loop}_n(\mathbb{S}_T)$. By the first statement, for all $a \in \Delta_{\mathfrak{l}}(T)$ and $l \in \mathfrak{l}(a)$,

$$\Phi_T(l) = \phi_{\mathfrak{l}}(a).$$

Hence Φ_T is invariant under area-preserving homeomorphisms. We take $t = 1$ in (44) and use the estimate (37) to see that

$$|\Phi_T(l) - \phi_T(n_*, a_0, a_*)| \leq 2C_{\mathfrak{l}}(T - a_{k_0} - a_{k_*}).$$

It remains to show that Φ_T satisfies the Makeenko–Migdal equations (10) on $\text{Loop}_n(\mathbb{S}_T)$. Let l be a regular loop with n self-intersections. Let $i \in \mathcal{I}$ and let $\theta : [0, \eta) \times \mathbb{S}_T \rightarrow \mathbb{S}_T$ be a Makeenko–Migdal flow at (l, v_i) . Write $a_{\theta}(t)$ for the face-area vector of $l(t) = \theta(t, l)$. Then

$$a_{\theta}(t) = a + t\Xi_i$$

so, by the argument leading to (44),

$$\mathbb{E}(\text{tr}(H_{l(t)})) = \mathbb{E}(\text{tr}(H_l)) + \int_0^t \mathbb{E}(\text{tr}(H_{l_i(s)})\text{tr}(H_{\hat{l}_i(s)}))ds.$$

By bounded convergence, on letting $N \rightarrow \infty$, we obtain

$$\Phi_T(l(t)) = \Phi_T(l) + \int_0^t \Phi_T(l_i(s))\Phi_T(\hat{l}_i(s))ds$$

as required.

Suppose finally that $\Psi : \text{Loop}_n(\mathbb{S}_T) \rightarrow \mathbb{C}$ is another function with the same properties. We have to show that $\Psi = \Phi_T$ on $\text{Loop}_n(\mathbb{S}_T)$. Let \mathfrak{l} be a combinatorial planar loop with at most n self-intersections. Then, for all $a \in \Delta_{\mathfrak{l}}(T)$, the set of embedded loops $\mathfrak{l}(a)$ is non-empty, and Ψ takes a constant value on $\mathfrak{l}(a)$. So there is a unique function

$$\psi_{\mathfrak{l}} : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{C}$$

such that $\psi_{\mathfrak{l}}(a) = \Psi(l)$ for all $l \in \mathfrak{l}(a)$. Let i be a self-intersection of \mathfrak{l} and let $l \in \mathfrak{l}(a)$. Then there exists a Makeenko–Migdal flow θ at (l, v_i) . Thus, for t sufficiently small, we have $\theta(t, l) \in \mathfrak{l}(a + t\Xi_i)$ and hence

$$\psi_{\mathfrak{l}}(a + t\Xi_i) = \Psi(\theta(t, l)).$$

Since Ψ satisfies the Makeenko–Migdal equations, it follows that $\psi_{\mathfrak{l}}$ has a directional derivative at a given by

$$\Xi_i \psi_{\mathfrak{l}}(a) = \left. \frac{d}{dt} \right|_{t=0} \Psi(\theta(t, l)) = \Psi(l_i) \Psi(\hat{l}_i) = \psi_{l_i}(a) \psi_{\hat{l}_i}(a)$$

where l_i, \hat{l}_i are the loops obtained by splitting l at v_i , and $\mathfrak{l}_i, \hat{\mathfrak{l}}_i$ are the associated combinatorial loops.

Given $a \in \Delta_I(T)$, define $a(t)$ as at (38). Then, by the argument leading to (39), for all $t \in [0, 1)$,

$$\psi_I(a(t)) = \psi_I(a) + \sum_{i \in \mathcal{I}} \alpha_i \int_0^t \psi_{l_i}(a(s)) \psi_{\hat{l}_i}(a(s)) ds.$$

Since Ψ satisfies (13), in the limit $t \rightarrow 1$, we have

$$|\psi_I(a(t)) - \phi_T(n_*, a_0, a_*)| \leq C_n(T - a_{k_0}(t) - a_{k_*}(t)) \rightarrow 0.$$

By the inductive hypothesis, since l_i and \hat{l}_i have no more than $n - 1$ self-intersections,

$$\psi_{l_i}(a(s)) = \phi_{l_i}(a(s)), \quad \psi_{\hat{l}_i}(a(s)) = \phi_{\hat{l}_i}(a(s))$$

for all $s \in [0, 1)$. Hence $\Psi(l) = \psi_I(a) = \phi_I(a) = \Phi_T(l)$, showing that $\Psi = \Phi_T$ on $\text{Loop}_n(\mathbb{S}_T)$, as required. Hence both statements hold for n and the induction proceeds. \square

5. Extension to loops of finite length

5.1. Some estimates for piecewise geodesic loops. Our aim in this section is to prove Proposition 2.7, which is the final step in the proof of our main result Theorem 2.2. We follow a line of argument adapted from [8], where estimates are given for Wilson loops in the plane, which are uniform in N . Instead of using explicit formulas for expectations of Wilson loops, the idea is to revisit certain estimates which were used in the construction of the Yang–Mills measure [41, Section 3.3], and to show that, when applied to suitable functions, these estimates are uniform in N . For clarity, we shall give a more detailed account of the argument of [8, Theorem 4.1], reproducing part of the proof of [41, Section 3.3].

Write $\text{Path}_*(\mathbb{S}_T)$ and $\text{Loop}_*(\mathbb{S}_T)$ for the sets of piecewise geodesic paths and loops in \mathbb{S}_T . Set $\kappa = \sqrt{\pi T}/2$ and note that κ is the length of a great circle between antipodal points in \mathbb{S}_T . For $\alpha \in \text{Path}(\mathbb{S}_T)$, write $n_0(\alpha)$ for the smallest integer such that $2^{-n_0(\alpha)}\ell(\alpha) < \kappa$. For $n \geq n_0(\alpha)$, we define $D_n(\alpha) \in \text{Path}_*(\mathbb{S}_T)$ by parametrizing α by $[0, 1]$ at constant speed and then interpolating the points $(\alpha(k2^{-n}) : k = 0, 1, \dots, 2^n)$ by geodesics. Then $D_n(\alpha) \rightarrow \alpha$ in length as $n \rightarrow \infty$, so $\text{Path}_*(\mathbb{S}_T)$ is dense in $\text{Path}(\mathbb{S}_T)$ for the topology of convergence in length with fixed endpoints. Note in particular that, when $\ell(\alpha) < \kappa$, we use the notation $D_0(\alpha)$ for the unique geodesic with the same endpoints as α . Define, for $\alpha \in \text{Loop}(\mathbb{S}_T)$,

$$\Psi_N(\alpha) = \sqrt{\mathbb{E}(\text{tr}(I - H_\alpha))} = \sqrt{1 - \Phi_T^N(\alpha)}, \quad \Phi_T^N(\alpha) = \mathbb{E}(\text{tr}(H_\alpha))$$

where $H = (H_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ is a Yang–Mills holonomy field in $U(N)$. By Lemma 3.5, there is a $K_1 > 0$ such that, for all N and all $a \in [0, T]$, for all simple loops α bounding a domain of area a ,

$$\Psi_N(\alpha) \leq K_1 \sqrt{a}. \quad (46)$$

The function Ψ_N inherits from H the following properties: for all $\alpha, \beta \in \text{Loop}(\mathbb{S}_T)$ with $\alpha \sim \beta$,

$$\Psi_N(\alpha) = \Psi_N(\alpha^{-1}), \quad \Psi_N(\alpha) = \Psi_N(\beta) \quad (47)$$

and, for all pairs of paths $\gamma_1, \gamma_2 \in \text{Path}(\mathbb{S}_T)$ which concatenate to form a loop,

$$\Psi_N(\gamma_1 \gamma_2) = \Psi_N(\gamma_2 \gamma_1). \quad (48)$$

Moreover, using the identity

$$2\Psi_N(\alpha\beta)^2 = \mathbb{E}(\text{tr}((H_\alpha - H_\beta^*)(H_\alpha^* - H_\beta)))$$

and Cauchy–Schwarz, we obtain

$$\Psi_N(\alpha\beta) \leq \Psi_N(\alpha) + \Psi_N(\beta) \quad (49)$$

whenever α and β have the same base point.

Lemma 5.1. *There is a function $\Psi : \text{Loop}_*(\mathbb{S}_T) \rightarrow [0, \infty)$ satisfying (46), (47), (48) and (49) and such that $\Psi_N(\alpha) \rightarrow \Psi(\alpha)$ as $N \rightarrow \infty$ for all $\alpha \in \text{Loop}_*(\mathbb{S}_T)$.*

Proof. Let us first argue that, for any $\alpha \in \text{Loop}_*(\mathbb{S}_T)$, there is an $n \geq 0$, a combinatorial loop \mathfrak{l} having at most n self-intersections, and a sequence $(\alpha_k : k \in \mathbb{N})$ in $\text{Loop}_n(\mathbb{S}_T)$ such that α_k is a drawing of \mathfrak{l} for all k and $\alpha_k \rightarrow \alpha$ in length as $k \rightarrow \infty$. A loop $\alpha \in \text{Loop}_*(\mathbb{S}_T)$ is a finite concatenation $\gamma_1 \dots \gamma_m$ of segments of great circles, each of which we may assume to have length less than κ . Set $n = m(m-1)/2$. Consider the parametrized family in $\text{Loop}_*(\mathbb{S}_T)$ obtained by small perturbations of the segment endpoints. Since any two distinct segments γ_i and γ_j either intersect in at most one point, or are contained in the same great circle, the set of loops in this family which are not in $\text{Loop}_n(\mathbb{S}_T)$ is of measure zero for a random choice of endpoints. Hence there exists a sequence $(\alpha_k : k \in \mathbb{N})$ in $\text{Loop}_n(\mathbb{S}_T)$ such that $\alpha_k \rightarrow \alpha$ with fixed endpoints.

Since the set of combinatorial planar loops with n self-intersections is finite, we can assume without loss of generality that there is a combinatorial planar loop \mathfrak{l} such that α_k is a drawing of \mathfrak{l} for all $k \geq 1$ and such that $a(\alpha_k)$ converges in $\overline{\Delta_{\mathfrak{l}}(T)}$ as $k \rightarrow \infty$, with limit a say. Consider the Wilson loop function $\phi_{\mathfrak{l}}^N : \Delta_{\mathfrak{l}}(T) \rightarrow \mathbb{C}$. By Proposition 2.6, we know that $\phi_{\mathfrak{l}}^N$ is uniformly continuous on $\Delta_{\mathfrak{l}}(T)$ for all N and $\phi_{\mathfrak{l}}^N \rightarrow \phi_{\mathfrak{l}}$ uniformly on $\Delta_{\mathfrak{l}}(T)$ as $N \rightarrow \infty$. Hence $\phi_{\mathfrak{l}}$ has a continuous extension to $\overline{\Delta_{\mathfrak{l}}(T)}$, which we denote also by $\phi_{\mathfrak{l}}$, and $\phi_{\mathfrak{l}}^N \rightarrow \phi_{\mathfrak{l}}$ uniformly on $\overline{\Delta_{\mathfrak{l}}(T)}$. Now, as $N \rightarrow \infty$,

$$\Phi_T^N(\alpha) = \phi_{\mathfrak{l}}^N(a) \rightarrow \phi_{\mathfrak{l}}(a)$$

so we can conclude that the the following limit is well-defined

$$\Phi_T(\alpha) = \lim_{N \rightarrow \infty} \Phi_T^N(\alpha).$$

Let us set

$$\Psi(\alpha) = \sqrt{1 - \Phi_T(\alpha)}.$$

Then $\Psi(\alpha) = \lim_{N \rightarrow \infty} \Psi_N(\alpha)$ and the properties (46), (47), (48), (49) extend to Ψ on taking the limit $N \rightarrow \infty$. \square

We note for later use a further inequality which follows from (47), (48), (49): for all $\alpha, \beta \in \text{Loop}_*(\mathbb{S}_T)$ having the same base point,

$$|\Psi(\alpha) - \Psi(\beta)| \leq \Psi(\alpha\beta^{-1}). \quad (50)$$

The following isoperimetric inequality is shown in [41, Lemma 3.3.5]: *there is a constant $K_2 \in [\kappa^{-1}, \infty)$ such that, for all $a \in [0, T]$ and all $\alpha \in P_*(\mathbb{S}_T)$ of length $\ell(\alpha) < K_2^{-1}$ and such that the loop $s = \alpha^{-1}D_0(\alpha)$ is simple, we have*

$$\sqrt{a} \leq K_2 \ell(\alpha)^{3/4} (\ell(\alpha) - \ell(D_0(\alpha)))^{1/4} \quad (51)$$

where a is the smaller of the areas of the connected components of $\mathbb{S}_T \setminus s^*$. The next proposition follows a line of argument similar to [41, Lemma 3.3.4] but reformulated in a simpler way and taking care to obtain a constant K independent of N .

Proposition 5.1 *There is a constant $K \in [\kappa^{-1}, \infty)$ such that, for all $N \in \mathbb{N}$, all $n \geq 0$ and all $\alpha \in \text{Path}(\mathbb{S}_T)$ with $2^{-n}\ell(\alpha) < K^{-1}$, we have*

$$\Psi_N(\alpha D_n(\alpha)^{-1}) \leq K \ell(\alpha)^{3/4} (\ell(\alpha) - \ell(D_n(\alpha)))^{1/4}.$$

Moreover the same estimate holds for Ψ whenever $\alpha \in \text{Path}_*(\mathbb{S}_T)$.

Proof. The argument relies only on the properties (46),(47),(48), (49) which hold for both Ψ_N and Ψ , and the continuity of Ψ_N on $\text{Loop}(\mathbb{S}_T)$, which allows us to reduce to the case $\alpha \in \text{Path}_*(\mathbb{S}_T)$. We will write it out for Ψ . Consider first the case where $\alpha \in \text{Path}_*(\mathbb{S}_T)$, and α is injective with $\ell(\alpha) < \kappa$. Then (see [41, Proposition 3.3.6] and Figure 2 therein) we can write α as the concatenation $\alpha_1 \dots \alpha_p$ of its excursions away from, or along, $D_0(\alpha)$ to obtain a lasso decomposition

$$\alpha D_0(\alpha)^{-1} \sim l_1 \dots l_p, \quad l_i = \gamma_i s_i \gamma_i^{-1}, \quad s_i = \alpha_i D_0(\alpha_i)^{-1}$$

where $s_i \in \text{Loop}_*(\mathbb{S}_T)$ and $\gamma_i \in \text{Path}_*(\mathbb{S}_T)$ for all i , and where either s_i is simple or $\alpha_i = D_0(\alpha_i)$, and such that

$$\ell(\alpha) = \ell(\alpha_1) + \dots + \ell(\alpha_p), \quad \ell(D_0(\alpha)) = \ell(D_0(\alpha_1)) + \dots + \ell(D_0(\alpha_p)).$$

Write a_i for the smaller of the areas of the connected components of $\mathbb{S}_T \setminus s_i^*$. In the case $\alpha_i = D_0(\alpha_i)$, when there is only one such component, set $a_i = 0$. Note that $\ell(s_i) \leq 2\ell(\alpha_i) \leq 2\ell(\alpha)$. Take $K = \max\{K_1 K_2, 2K_2\}$ and suppose that $\ell(\alpha) < K^{-1}$. Then

$$\begin{aligned} \Psi(\alpha D_0(\alpha)^{-1}) &= \Psi(l_1 \dots l_p) \leq \Psi(l_1) + \dots + \Psi(l_p) = \Psi(s_1) + \dots + \Psi(s_p) \\ &\leq K_1(\sqrt{a_1} + \dots + \sqrt{a_p}) \\ &\leq K_1 K_2 \sum_i \ell(\alpha_i)^{3/4} (\ell(\alpha_i) - \ell(D_0(\alpha_i)))^{1/4} \\ &\leq K \ell(\alpha)^{3/4} (\ell(\alpha) - \ell(D_0(\alpha)))^{1/4} \end{aligned}$$

where we used Hölder's inequality for the last step.

Now, for general $\alpha \in \text{Path}_*(\mathbb{S}_T)$ with $\ell(\alpha) < K^{-1}$, according to [41, Proposition 1.4.9], we can write α as a concatenation $\alpha_1 \dots \alpha_p \gamma$, where each α_i is run up to the end of its first interval of self intersection. So we obtain a lasso decomposition

$$\alpha \sim l_1 \dots l_p \gamma, \quad l_i = \gamma_i s_i \gamma_i^{-1}, \quad \ell(\alpha) = \ell(s_1) + \dots + \ell(s_p) + \ell(\gamma)$$

where $s_i \in \text{Loop}_*(\mathbb{S}_T)$ is simple and $\gamma_i \in \text{Path}_*(\mathbb{S}_T)$ for all i , and where $\gamma \in \text{Path}_*(\mathbb{S}_T)$ is injective. Write a_i for the smaller of the areas of the connected components of $\mathbb{S}_T \setminus s_i^*$. Then

$$\Psi(l_i) = \Psi(s_i) \leq K_1 \sqrt{a_i} \leq K_1 K_2 \ell(s_i)$$

so

$$\Psi(l_1) + \dots + \Psi(l_p) \leq K(\ell(s_1) + \dots + \ell(s_p)) = K(\ell(\alpha) - \ell(\gamma)).$$

On the other hand, by the first part,

$$\Psi(\gamma D_0(\gamma)^{-1}) \leq K \ell(\gamma)^{3/4} (\ell(\gamma) - \ell(D_0(\gamma)))^{1/4}.$$

But $D_0(\gamma) = D_0(\alpha)$, so

$$\begin{aligned} \Psi(\alpha D_0(\alpha)^{-1}) &= \Psi(l_1 \dots l_p \gamma D_0(\gamma)^{-1}) \leq \Psi(l_1) + \dots + \Psi(l_p) + \Psi(\gamma D_0(\gamma)^{-1}) \\ &\leq K(\ell(\alpha) - \ell(\gamma)) + K \ell(\gamma)^{3/4} (\ell(\gamma) - \ell(D_0(\gamma)))^{1/4} \\ &\leq K \ell(\alpha)^{3/4} (\ell(\alpha) - \ell(D_0(\alpha)))^{1/4}. \end{aligned}$$

Finally, for $n \geq 0$ and $\alpha \in \text{Path}_*(\mathbb{S}_T)$ with $2^{-n} \ell(\alpha) < K^{-1}$, we can write α as a concatenation $\alpha_1 \dots \alpha_{2^n}$ such that

$$D_n(\alpha) = D_0(\alpha_1) \dots D_0(\alpha_{2^n}), \quad \ell(\alpha_i) = 2^{-n} \ell(\alpha).$$

Then there is a lasso decomposition

$$\alpha D_n(\alpha)^{-1} \sim l_1 \dots l_{2^n}, \quad l_i = \gamma_i \alpha_i D_0(\alpha_i)^{-1} \gamma_i^{-1}$$

where $\gamma_i \in \text{Loop}_*(\mathbb{S}_T)$ for all i . Then, by the second part,

$$\begin{aligned} \Psi(\alpha D_n(\alpha)^{-1}) &= \Psi(l_1 \dots l_{2^n}) \leq \sum_i \Psi(\alpha_i D_0(\alpha_i)^{-1}) \\ &\leq \sum_i K \ell(\alpha_i)^{3/4} (\ell(\alpha_i) - \ell(D_0(\alpha_i)))^{1/4} \\ &\leq K \ell(\alpha)^{3/4} (\ell(\alpha) - \ell(D_n(\alpha)))^{1/4}. \end{aligned}$$

□

5.2. Proof of Proposition 2.7. Let $(H_\gamma : \gamma \in \text{Path}(\mathbb{S}_T))$ be a Yang–Mills holonomy field in $U(N)$. We have to show that $\text{tr}(H_l)$ converges in probability as $N \rightarrow \infty$ for all $l \in \text{Loop}(\mathbb{S}_T)$. Further, we have to show that the master field

$$\Phi_T(l) = \lim_{N \rightarrow \infty} \mathbb{E}(\text{tr}(H_l))$$

is the unique continuous function $\text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$ which is invariant under reduction and under area-preserving homeomorphisms, satisfies the Makeenko–Migdal equations (10) on regular loops, and satisfies (11) for simple loops.

Let $l \in \text{Loop}(\mathbb{S}_T)$ and set $l_n = D_n(l)$. Note that $D_n(l_m) = l_n$ when $m \geq n$. By (50) and Proposition 5.1, for $n \in \mathbb{N}$ sufficiently large and $m \geq n$,

$$|\Psi(l_m) - \Psi(l_n)| \leq \Psi(l_m l_n^{-1}) \leq K \ell(l)^{3/4} (\ell(l) - \ell(l_n))^{1/4}.$$

Also

$$\mathbb{E}(|\text{tr}(H_{l_n}) - \text{tr}(H_l)|^2) \leq \mathbb{E}(\text{tr}((H_{l_n} - H_l)(H_{l_n} - H_l)^*)) = 2\Psi_N(l l_n^{-1})^2$$

so

$$\|\text{tr}(H_{l_n}) - \text{tr}(H_l)\|_2 \leq \sqrt{2} \Psi_N(l l_n^{-1}) \leq \sqrt{2} K \ell(l)^{3/4} (\ell(l) - \ell(l_n))^{1/4}.$$

Since $\ell(l_n) \rightarrow \ell(l)$ as $n \rightarrow \infty$, we see that $\Psi(l_n)$ and $\Phi_T(l_n) = 1 - \Psi(l_n)^2$ must converge as $n \rightarrow \infty$. Define

$$\tilde{\Phi}(l) = \lim_{n \rightarrow \infty} \Phi_T(l_n).$$

Let $n \rightarrow \infty$ and then $N \rightarrow \infty$ in the inequality

$$\|\text{tr}(H_l) - \tilde{\Phi}(l)\|_1 \leq \|\text{tr}(H_l) - \text{tr}(H_{l_n})\|_1 + \|\text{tr}(H_{l_n}) - \Phi_T(l_n)\|_1 + |\Phi_T(l_n) - \tilde{\Phi}(l)|$$

to see that $\text{tr}(H_l) \rightarrow \tilde{\Phi}(l)$ in probability and $\Phi_T(l) = \lim_{N \rightarrow \infty} \mathbb{E}(\text{tr}(H_l)) = \tilde{\Phi}(l)$.

The invariance of Φ_T on $\text{Loop}(\mathbb{S}_T)$ under reduction and area-preserving homeomorphisms follows from the corresponding invariance properties of Φ_T^N . The claimed properties of Φ_T on simple and regular loops were shown in Propositions 2.5 and 2.6. We now show that Φ_T is continuous on $\text{Loop}(\mathbb{S}_T)$. For this, we translate to our context the argument of [41, Proposition 3.3.9]. Let $\alpha \in \text{Loop}(\mathbb{S}_T)$ and let $(\alpha_n : n \in \mathbb{N})$ be a sequence in $\text{Loop}(\mathbb{S}_T)$ which converges to α in length. We have to show that $\Phi_T(\alpha_n) \rightarrow \Phi_T(\alpha)$. There exist area-preserving homeomorphisms θ_n on \mathbb{S}_T such that $\theta_n(\alpha_n)$ converges to α in length with fixed endpoints. We have $\Phi_T(\alpha) = 1 - \Psi(\alpha)^2$ and we know that $\Psi(D_m(\alpha_n)) \rightarrow \Psi(\alpha_n)$ as $m \rightarrow \infty$. Hence it will suffice to consider the case where α_n is piecewise geodesic for all n and α_n converges to α in length with fixed endpoints,

and to show then that $\Psi(\alpha_n) \rightarrow \Psi(\alpha)$ as $n \rightarrow \infty$. Parametrize α at constant speed and choose parametrizations for the loops α_n so that

$$\|\alpha_n - \alpha\|_\infty = \sup_{t \in [0,1]} |\alpha_n(t) - \alpha(t)| \rightarrow 0.$$

Fix $m \geq 0$ and write $D_m(\alpha)$ and α_n as concatenations

$$D_m(\alpha) = \sigma_1 \dots \sigma_{2^m}, \quad \alpha_n = \alpha_{n,1} \dots \alpha_{n,2^m}$$

where σ_i is the geodesic from $\alpha((i-1)2^{-m})$ to $\alpha(i2^{-m})$ and $\alpha_{n,i}$ is the restriction of α_n to $[(i-1)2^{-m}, i2^{-m}]$. For $i = 0, 1, \dots, 2^m$, denote by $\eta_{n,i}$ the geodesic from $\alpha(i2^{-m})$ to $\alpha_n(i2^{-m})$. Then $\ell(\eta_{n,0}) = \ell(\eta_{n,2^m}) = 0$ and, for $i = 1, \dots, 2^m - 1$,

$$\ell(\eta_{n,i}) \leq \|\alpha_n - \alpha\|_\infty.$$

Set

$$\beta_n = \beta_{n,1} \dots \beta_{n,2^m}, \quad \beta_{n,i} = \eta_{n,i-1} \alpha_{n,i} \eta_{n,i}^{-1}.$$

Then $\alpha_n \sim \beta_n$ and $D_0(\beta_{n,i}) = \sigma_i$ for all i . So

$$\Psi(\alpha_n D_m(\alpha)^{-1}) = \Psi(\beta_n D_m(\alpha)^{-1})$$

and, by the argument used in the last part of the proof of Proposition 5.1,

$$\Psi(\beta_n D_m(\alpha)^{-1}) \leq K \ell(\beta_n)^{3/4} (\ell(\beta_n) - \ell(D_m(\alpha)))^{1/4}.$$

Now

$$\ell(\beta_n) \leq \ell(\alpha_n) + 2^{m+1} \|\alpha_n - \alpha\|_\infty$$

so

$$\begin{aligned} |\Psi(\alpha_n) - \Psi(\alpha)| &\leq \Psi(\alpha_n D_m(\alpha)^{-1}) + |\Psi(D_m(\alpha)) - \Psi(\alpha)| \\ &\leq K(\ell(\alpha_n) + 2^{m+1} \|\alpha_n - \alpha\|_\infty)^{3/4} (\ell(\alpha_n) - \ell(D_m(\alpha)) + 2^{m+1} \|\alpha_n - \alpha\|_\infty)^{1/4} \\ &\quad + |\Psi(D_m(\alpha)) - \Psi(\alpha)|. \end{aligned}$$

On letting first $n \rightarrow \infty$ and then $m \rightarrow \infty$, we see that $\Psi(\alpha_n) \rightarrow \Psi(\alpha)$ as required.

Finally, suppose that $\Psi : \text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$ is another function with the same properties. Then $\Psi = \Phi_T$ on $\text{Loop}_0(\mathbb{S}_T)$. Suppose inductively for $n \geq 1$, that $\Psi = \Phi_T$ on $\text{Loop}_{n-1}(\mathbb{S}_T)$, and let $l \in \text{Loop}_n(\mathbb{S}_T)$. We follow the argument at the end of the proof of Proposition 2.6, except that, in place of the estimate (13), we use the cylinder-based loop deformation l_t defined at (42) and the continuity of Ψ to see that

$$\psi_l(a(t)) = \Psi(l_t) \rightarrow \Psi(l_1) = \Psi(s^{n*}) = \phi_T(n_*, a_0, a_*).$$

Then we can conclude as before that $\Psi = \Phi_T$ on $\text{Loop}_n(\mathbb{S}_T)$. Hence, by induction, $\Psi = \Phi_T$ on all regular loops. But these are dense in $\text{Loop}(\mathbb{S}_T)$ and Ψ and Φ_T are continuous, so $\Psi = \Phi_T$ on $\text{Loop}(\mathbb{S}_T)$. \square

6. Further properties of the master field

6.1. Relation with the Hermitian Brownian loop. Let $W = (W_t : t \geq 0)$ be a Brownian motion in the set of $N \times N$ Hermitian matrices $H(N)$ equipped with the inner product

$$\langle h_1, h_2 \rangle = N \operatorname{Tr}(h_1 h_2^*).$$

Let $w = (w_t : t \geq 0)$ be a free Brownian motion, defined on some non-commutative probability space (\mathcal{A}, τ) . The inner product is scaled with N so that W converges in non-commutative distribution (in probability) to w , that is to say, for all $n \in \mathbb{N}$ and all $t_1, \dots, t_n \geq 0$,

$$\operatorname{tr}(W_{t_1} \dots W_{t_n}) \rightarrow \tau(w_{t_1} \dots w_{t_n})$$

in probability as $N \rightarrow \infty$. Fix $T \in (0, \infty)$ and define for $t \in [0, T]$

$$B_t = W_t - \frac{t}{T} W_T, \quad b_t^T = w_t - \frac{t}{T} w_T.$$

Then $B = (B_t : t \in [0, T])$ is a Hermitian Brownian loop in $H(N)$ and B converges in non-commutative distribution to $b^T = (b_t^T : t \in [0, T])$. The non-commutative process b^T is called the *free Hermitian Brownian loop*. We will write simply b for b^1 .

Let $x = (x_t : t \in [0, T])$ be a free unitary Brownian loop in (\mathcal{A}, τ) , as defined in Subsection 2.7.

Proposition 6.1 *Suppose that $T \in (0, \pi^2]$. Then, for all $t \in [0, T]$ and all $n \in \mathbb{Z}$,*

$$\tau(x_t^n) = \int_{\mathbb{R}} e^{inx} s_{\sqrt{t(T-t)/T}}(x) dx = \tau(e^{inb_t^T})$$

where s_t is the semi-circle density (15) of variance t . On the other hand, for almost all T and almost all $s, t \in (0, T)$ with $s < t$,

$$\tau(x_s^* x_t) \neq \tau(e^{-ib_s^T} e^{ib_t^T})$$

so $(e^{ib_t^T} : t \in [0, T])$ is not a free unitary Brownian loop.

Proof. The first assertion is the content of Proposition 3.6. We turn to the second assertion. Let $(X_t : t \in [0, T])$ be a Brownian loop in $U(N)$ based at 1. Then, since Brownian motion in $U(N)$ is a Lévy process, $X_s^{-1} X_t$ has same law as X_{t-s} . On letting $N \rightarrow \infty$, we deduce that

$$\tau(x_s^* x_t) = \tau(x_{t-s}) = \tau(e^{ib_{t-s}^T}) = \tau(e^{i(b_t^T - b_s^T)})$$

where we used free independence and stationarity of the increments of free Brownian motion for the last equality. Hence, by the scaling properties of free Brownian motion,

$$\tau(e^{-ib_s^T} e^{ib_t^T}) - \tau(x_s^* x_t) = \tau(e^{-ib_s^T} e^{ib_t^T} - e^{i(b_t^T - b_s^T)}) = F_{s/T, t/T}(\sqrt{T})$$

where

$$F_{s,t}(\sigma) = \tau(e^{-i\sigma b_s} e^{i\sigma b_t} - e^{i\sigma(b_t - b_s)}).$$

By Fubini's theorem, it will suffice to show, for all $s, t \in (0, 1)$ with $s < t$, that $F_{s,t}(\sigma) \neq 0$ for almost all $\sigma \in (0, \pi]$. We expand the exponential function up to fourth order and use scale invariance of free Brownian motion to obtain

$$F_{s,t}(0) = F'_{s,t}(0) = F''_{s,t}(0) = F'''_{s,t}(0) = 0, \quad F''''_{s,t}(0) = 2\tau(b_s^2 b_t^2 - b_s b_t b_s b_t).$$

The variables $(b_t : t \in [0, 1])$ are semi-circular, therefore all free cumulants of order more than 3 vanish (see for example [48, equation 11.4]). So, using the decomposition of moments into free cumulants⁶ (see [48, equation 11.8]),

$$\tau(b_s^2 b_t^2 - b_s b_t b_s b_t) = \tau(b_s^2) \tau(b_t^2) - \tau(b_s b_t)^2 = s(t-s)(1-t) > 0.$$

Since $F_{s,t}$ is analytic in σ on $(0, \pi]$, this implies that it has at most finitely many zeros. \square

⁶ Here it can be understood as a ‘non-commutative’ Wick formula, with non-crossing matchings in place of all matchings.

6.2. Duality at the midpoint of the loop. Recall from (6) and (8) the form of ρ_T . It will be convenient to set $\alpha = 0$ and $\beta = 2/\sqrt{T}$ in the subcritical case $T \in (0, \pi^2]$. Denote by \mathbb{H} the open upper half-plane. The following relation appeared first in the physics literature [28, equation 1.2], without a mathematical proof.

Proposition 6.2 *Let $(x_t : t \in [0, T])$ be a free unitary Brownian loop. Then, for all $T > 0$, the spectral measure of $x_{T/2}$ has a density ρ_T^* with respect to Lebesgue measure on \mathbb{U} (of mass 2π), which is invariant under complex conjugation and is such that*

$$\pi\rho_T^* : \mathbb{U} \cap \mathbb{H} \rightarrow (\alpha, \beta)$$

is the inverse mapping of

$$e^{i\pi\rho_T} : (\alpha, \beta) \rightarrow \mathbb{U} \cap \mathbb{H}.$$

Proof. We write the proof for the supercritical case $T > \pi^2$, leaving the minor adjustments needed when $T \leq \pi^2$ to the reader. The function $\rho_T : (\alpha, \beta) \rightarrow (0, 1)$ is continuous and strictly decreasing, with $\rho_T(\alpha) = 1$ and $\rho_T(\beta) = 0$. Indeed, according to formula (8) and an elementary computation (see for example [45, equation 150]), for $x \in (\alpha, \beta)$,

$$\frac{\pi\alpha}{2} \sqrt{(x^2 - \alpha^2)(\beta^2 - x^2)} \rho_T'(x) = \int_0^1 \frac{\alpha^2 s^2 - x^2}{\beta^2 \sqrt{(1 - s^2)(1 - k^2 s^2)}} ds < 0.$$

Write ψ for the inverse of the bijection $\pi\rho_T : (\alpha, \beta) \rightarrow (0, \pi)$. For all $n \in \mathbb{Z} \setminus \{0\}$, by Lemma 3.7,

$$\tau(x_{T/2}^n) = \frac{2}{n\pi} \int_\alpha^\beta \sin\{n\pi\rho_T(x)\} dx = -\frac{2}{n\pi} \int_0^\pi \sin(n\theta) \psi'(\theta) d\theta.$$

We integrate by parts to obtain

$$\tau(x_{T/2}^n) = \frac{2}{\pi} \int_0^\pi \cos(n\theta) \psi(\theta) d\theta.$$

Hence the spectral measure of $x_{T/2}$ has a density ρ_T^* with respect to Lebesgue measure on \mathbb{U} given by

$$\rho_T^*(e^{i\theta}) = \psi(|\theta|)/\pi, \quad |\theta| \leq \pi.$$

□

6.3. Convergence to the planar master field. We now investigate the behaviour of the master field Φ_T as $T \rightarrow \infty$. For $T > 0$, $n \in \mathbb{N}$ and $t \in [0, T]$, set

$$m_T(n, t) = \Phi_T(l^n)$$

where l is a simple loop which divides \mathbb{S}_T into components of areas t and $T - t$. Recall that $m_T(n, t)$ does not depend on the choice of l .

Proposition 6.3 *We have*

$$m_T(n, t) \rightarrow \frac{e^{-nt/2}}{2\pi i n} \int_\gamma \left(1 + \frac{1}{z}\right)^n e^{-ntz} dz = e^{-nt/2} \sum_{k=1}^{n-1} \frac{(-t)^k}{k!} \binom{n}{k+1} n^{k-1}$$

uniformly in $t \in [0, T]$ as $T \rightarrow \infty$, where γ is any positively oriented loop in \mathbb{C} winding once around 0.

Proof. Since the second complete elliptic integral $E(k)$ is bounded and the first $K(k)$ is bounded on compacts in $[0, 1)$, the relation

$$T = 8EK - 4(1 - k^2)K^2$$

forces $k \rightarrow 1$ as $T \rightarrow \infty$. Since $\alpha = k\beta \leq 1/2$ and $\beta \geq 1/2$ for all T , this implies $\alpha, \beta \rightarrow 1/2$ as $T \rightarrow \infty$. Hence $\rho_T(x) \rightarrow 1$ for $|x| < 1/2$ and $\rho_T(x) \rightarrow 0$ for $|x| > 1/2$, so

$$G_T(z) = \int_{\mathbb{R}} \frac{\rho_T(x)}{z - x} dx \rightarrow \int_{-1/2}^{1/2} \frac{dx}{z - x} = \text{Log} \left(\frac{z + 1/2}{z - 1/2} \right)$$

uniformly on compacts in $\{z \in \mathbb{C} : |z| > 1/2\}$. By Proposition 3.7, for $R > 1/2$ and T sufficiently large, uniformly in $t \in [0, T]$,

$$\begin{aligned} m_T(n, t) &= \frac{1}{2\pi i n} \int_{\gamma_R} \exp\{-n(tz - G_T(z))\} dz \\ &\rightarrow \frac{1}{2\pi i n} \int_{\gamma_R} e^{-ntz} \left(\frac{z + 1/2}{z - 1/2} \right)^n dz = \frac{e^{-nt/2}}{2\pi i n} \int_{\gamma} \left(1 + \frac{1}{z} \right)^n e^{-ntz} dz \end{aligned}$$

where γ_R is the positively oriented boundary of $\{z \in \mathbb{C} : |z| = R\}$. \square

We remark that it is known from Biane [4] that the free unitary Brownian motion has spectral measure ν_t on the circle \mathbb{U} , with moments given for all $n \in \mathbb{Z}$, by

$$\int_{\mathbb{U}} \omega^n \nu_t(d\omega) = e^{-|n|t/2} \sum_{k=1}^{|n|-1} \frac{(-t)^k}{k!} \binom{|n|}{k+1} |n|^{k-1}$$

whereas the spectral measure ν_t^N of a Brownian motion on $U(N)$ at time t satisfies

$$\nu_t^N \longrightarrow \nu_t$$

weakly in probability on \mathbb{U} as $N \rightarrow \infty$. See also Lévy [40], for another proof using Schur–Weyl duality. Let us write $\nu_{t,T}^N$ and $\nu_{t,T}$ for the spectral measure of a marginal of respectively the unitary Brownian loop on $U(N)$ and of the free unitary Brownian loop (defined in Section 2.7), where both processes have lifetime T and the marginal is taken at time t . We can now deduce that the following diagram commutes

$$\begin{array}{ccc} \nu_{t,T}^N & \xrightarrow{N \rightarrow \infty} & \nu_{t,T} \\ T \rightarrow \infty \downarrow & & \downarrow T \rightarrow \infty \\ \nu_t^N & \xrightarrow{N \rightarrow \infty} & \nu_t \end{array}$$

where the horizontal arrows denote weak limits on \mathbb{U} as $N \rightarrow \infty$, which follow from Theorem 2.9 and Corollary 2.8, along with [4], whereas the left vertical arrow denotes a limit in law on $\mathcal{M}_1(\mathbb{U})$ and the right vertical arrow denotes the weak limit induced by Proposition 6.3. The next proposition enables us to show that such a commutative diagram holds true for the non-commutative distributions. We will not give details here on this latter point.

Denote by $\text{Loop}(\mathbb{R}^2)$ the set of loops of finite length in \mathbb{R}^2 and let

$$\Phi : \text{Loop}(\mathbb{R}^2) \rightarrow [-1, 1]$$

be the planar master field as defined in [42].

Proposition 6.4 *For each $T > 0$, fix a point $y_T \in \mathbb{S}_T$ and denote by p_T the inverse map of the stereographic projection $\mathbb{S}_T \setminus \{y_T\} \rightarrow \mathbb{C}$. Then, for all $l \in \text{Loop}(\mathbb{R}^2)$,*

$$\Phi_T(p_T(l)) \rightarrow \Phi(l) \quad \text{as } T \rightarrow \infty.$$

Proof. Let l be a simple loop in $\text{Loop}(\mathbb{R}^2)$ and denote by a the finite area enclosed by l . Then $p_T(l)$ is a simple loop in $\text{Loop}(\mathbb{S}_T)$ which divides \mathbb{S}_T into two components and does not pass through y_T . Denote by a_T the area of the component which does not contain y_T . Then $a_T \rightarrow a$ as $T \rightarrow \infty$. By Proposition 6.3, this implies

$$\Phi_T(p_T(l^n)) \rightarrow e^{-na/2} \sum_{k=1}^{n-1} \frac{(-a)^k}{k!} \binom{n}{k+1} n^{k-1} = \Phi(l^n)$$

as $T \rightarrow \infty$, where we used [42, equation (2)] for the last equality.

Now Φ also satisfies the Makeenko–Migdal equations [42]. By a variation of the argument used to prove Theorem 2.6, we can extend convergence from powers of simple loops to all regular loops. We sketch the small change which is needed. There is now a face, k_∞ say, of infinite area. So we work in the orthant

$$Y_l = \{(a_k : k \in \mathcal{F}_l) : a_{k_\infty} = \infty \text{ and } a_k \in (0, \infty) \text{ for all } k \neq k_\infty\}.$$

Set

$$\bar{a} = \sum_{k \neq k_\infty} a_k.$$

Write k_0, k_* , as before, for the faces of minimal and maximal winding number, now choosing the additive constant so that $n_l(k_\infty) = 0$. Given $a \in Y_l$, either $\langle a, n_l \rangle \geq 0$, or $\langle a, n_l \rangle < 0$. (We use here the convention that $\infty \times 0 = 0$.) In the first case, $k_* \neq k_\infty$ and there exists $a' \in \bar{Y}_l$ with $a'_k = 0$ for $k \neq k_*, k_\infty$ such that

$$\langle a', n_l \rangle = a'_{k_*} n_l(k_*) = \langle a, n_l \rangle.$$

Set

$$v_k = \begin{cases} a'_k - a_k, & \text{if } k \neq k_\infty, \\ \bar{a} - \bar{a}', & \text{if } k = k_\infty. \end{cases}$$

and set $a(t) = a + tv$. Then $a' = a(1)$ and $a(t) \in Y_l$ for all $t \in [0, 1]$, and $v \in \mathfrak{m}_l$ by Proposition 4.5. An analogous argument holds in the second case. We can then proceed as in Subsection 4.5. The arguments of Section 5 also carry over to extend the limit

$$\Phi_T(p_T(l)) \rightarrow \Phi(l)$$

to all $l \in \text{Loop}(\mathbb{R}^2)$. We omit the details. \square

6.4. Uniqueness of the master field. In Theorem 2.4, we showed that the master field is characterized by certain properties. In fact there is some redundancy in this characterization, as the following result shows in replacing property (11) by (52), which is the case $n = 1$ of (11).

Proposition 6.5 *Let $\Phi : \text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$ be a continuous function, which is invariant under reduction and under area-preserving, orientation-preserving homeomorphisms, satisfies the Makeenko–Migdal equations on regular loops, and is given on simple loops l by*

$$\Phi(l) = \frac{2}{\pi} \int_0^\infty \cosh\{(a-b)x/2\} \sin\{\pi\rho_T(x)\} dx \quad (52)$$

where a and b are the areas of the connected components of $\mathbb{S}_T \setminus \{l^\}$. Then Φ is the master field Φ_T .*

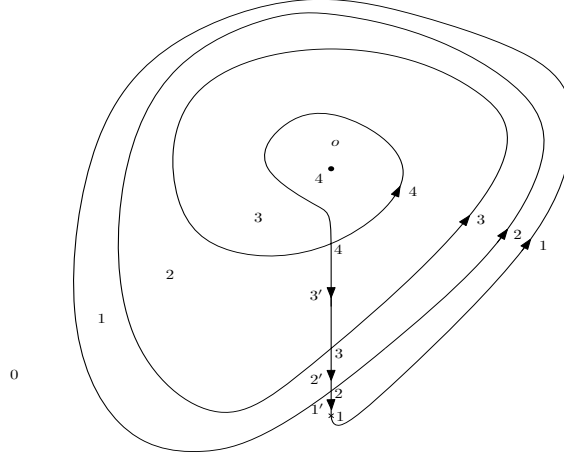


Fig. 6. A drawing of the maximally winding loop \mathfrak{l}_4 .

Furthermore, we shall see in the next section that given the value (52) of $\Phi_T(l)$ for any simple loops, it is possible to obtain new explicit formulas for $\Phi_T(l^n)$ using the Makeenko-Migdal equations (see (55)), matching with (11). The proof of proposition 6.5 will be based on an argument for a special class of loops which we now introduce. Informally, for $n \geq 1$ fix an initial point x_1 and draw an inward anticlockwise spiral which winds n times around another point o , crossing the line ox_1 at points x_2, \dots, x_n then, on hitting ox_1 for the n th time, returning to x_1 along ox_1 . Thus we obtain a combinatorial planar loop \mathfrak{l}_n whose combinatorial graph is given as follows:

$$\mathcal{V} = \{1, \dots, n\}, \quad \mathcal{E} = \{1, \dots, n\} \cup \{1', \dots, (n-1)'\}, \quad \mathcal{F} = \{0, 1, \dots, n\}$$

where, for $j = 1, \dots, n-1$,

$$s(j) = j, \quad t(j) = j+1, \quad s(j') = j+1, \quad t(j') = j$$

and

$$l(j) = l(j') = j, \quad r(j) = r(j') = j-1$$

while

$$s(n) = t(n) = n, \quad l(n) = n, \quad r(n) = n-1.$$

See Figure 6. Here, we have used a non-standard labelling for the edges and faces which is adapted to the structure of the graph. Note that the self-intersections of \mathfrak{l}_n are labelled by $\{2, \dots, n\}$. If we fix the additive constant for the winding number so that $n_{\mathfrak{l}_n}(0) = 0$, then $n_{\mathfrak{l}_n}(n) = n$. For $n \geq 1$ and for any combinatorial planar loop \mathfrak{l} with $n-1$ self-intersections, we have

$$n_* = \max\{n_{\mathfrak{l}}(k) - n_{\mathfrak{l}}(k') : k, k' \in \mathcal{F}\} \leq n.$$

We call \mathfrak{l}_n , and any associated regular embedded loop l , and any rerooting of l , a *maximally winding* loop.

Proof of Proposition 6.5. Let $n \geq 1$ and suppose inductively that $\Phi(l^m) = \Phi_T(l^m)$ for all $m \leq n$ and all simple loops l . A comparison of equations (11) and (52) shows that this is true for $n = 1$. Let l be a simple loop which divides \mathbb{S}_T into components of areas $a_0, a_* \in (0, T)$. We can find $(\alpha_1, \alpha_2, \dots, \alpha_{n+2})$ such that

$$\alpha_1 = 0, \quad \alpha_2 = a_0, \quad \alpha_{n+1} = a_*, \quad \alpha_{n+2} = 0$$

and, for $m = 2, \dots, n+1$,

$$\alpha_{m-1} - 2\alpha_m + \alpha_{m+1} < 0.$$

Consider the (constant) vector field v on $\Delta_{n+1}(T)$ given by

$$v = \sum_{i=2}^{n+1} \alpha_i \Xi_i.$$

Then $v_0 = -a_0$ and $v_{n+1} = -a_*$ and $v_k > 0$ for $k = 1, \dots, n$. Set

$$a(t) = (a_0, 0, \dots, 0, a_*) + tv$$

then $a(t) \in \Delta_{n+1}(T)$ for all $t \in (0, 1)$ and $a_0(1) = a_{n+1}(1) = 0$. There exists a continuous family of loops $(l(t) : t \in [0, 1])$, with a common basepoint such that, $l(0) = l^{n+1}$, $l(t) \in \mathcal{G}_{n+1}(a(t))$ for all $t \in (0, 1)$, and $l(1)$ is a maximally winding loop with $n-2$ self-intersections. Then, by the arguments used in the proof of Proposition 2.6,

$$\Phi(l(1)) = \Phi(l^{n+1}) + \sum_{i=2}^{n+1} \alpha_i \int_{\tau}^1 \Phi(l_i(s)) \Phi(\hat{l}_i(s)) ds$$

where $l_i(s)$ and $\hat{l}_i(s)$ are maximally winding loops having $i-2$ and $n+1-i$ self-intersections. But the same equation holds for Φ_T and the inductive hypothesis, combined with the argument of the proof of Proposition 2.6, implies that

$$\Phi(l(1)) = \Phi_T(l(1)), \quad \Phi(l_i(s)) = \Phi_T(l_i(s)), \quad \Phi(\hat{l}_i(s)) = \Phi_T(\hat{l}_i(s)).$$

Hence $\Phi(l^{n+1}) = \Phi_T(l^{n+1})$ and the induction proceeds. Finally, by Proposition 2.6, it follows that $\Phi(l) = \Phi_T(l)$ for all $l \in \text{Loop}(\mathbb{S}_T)$. \square

On the other hand, condition (52) is not redundant in Proposition 6.5, as we now show. Each loop $l \in \text{Loop}(\mathbb{S}_T)$ has a winding number function

$$n_l : \mathbb{S}_T \setminus l^* \rightarrow \mathbb{Z}$$

which is unique up to an additive constant. By the Banchoff–Pohl inequality [2], we know that $n_l \in L^2(\mathbb{S}_T)$ so n_l has a well-defined average value $\langle n_l \rangle$ with respect to the uniform distribution on \mathbb{S}_T , up to the same additive constant. Hence, we can define a unique function $\Psi : \text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$ by

$$\Psi(l) = e^{2\pi i \langle n_l \rangle}.$$

For loops l_1, l_2 based at the same point, we have $n_{l_1 l_2} = n_{l_1} + n_{l_2}$, so

$$\Psi(l_1 l_2) = \Psi(l_1) \Psi(l_2).$$

Moreover, Ψ is invariant under any area-preserving, orientation-preserving diffeomorphism so, in particular, under any Makeenko–Migdal flow. Consider, for $n \in \mathbb{Z}$, the twisted master field $\Phi_T^{(n)} : \text{Loop}(\mathbb{S}_T) \rightarrow \mathbb{C}$ given by

$$\Phi_T^{(n)}(l) = \Psi(l)^n \Phi_T(l).$$

Then $\Phi_T^{(n)}$ is continuous, invariant under reduction and area-preserving, orientation-preserving homeomorphisms and satisfies the Makeenko–Migdal equations on regular loops. However, for a simple loop l which winds positively around a domain of area a , we have

$$\Psi(l) = e^{2\pi i a/T}$$

so, for $n \neq 0$, $\Phi_T^{(n)}$ is not the master field. Hence, by Proposition 6.5, or by inspection, $\Phi_T^{(n)}$ does not satisfy (52).

For $n \neq 0$, the twisted field $\Phi_T^{(n)}$ from the preceding paragraph also fails to be invariant under orientation-reversing homeomorphisms. We do not know whether this stronger invariance condition would allow one to dispense with (52) in Proposition 6.5.

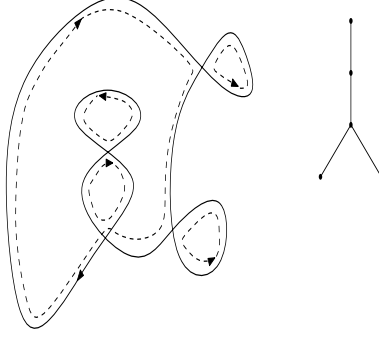


Fig. 7. A splittable combinatorial planar loop with its family of simple loops \mathcal{S}_l drawn in dashed lines, next to a combinatorial representation of the tree structure of \mathcal{S}_l .

6.5. Combinatorial formulas for the master field. Rusakov [49] proposed, without proof, that there should be a closed formula for the value of the master field for any regular loop on the sphere. We now prove a formula for a restricted class of loops introduced in [35], which agrees with (42) in [49]. We were not able to prove or disprove this latter formula in the general case⁷ and leave this question open. Let us say that a combinatorial planar loop l is *splittable*⁸ if for all self-intersection points i of l , the two loops l_i, \hat{l}_i , obtained by following outgoing strands of l starting from i , intersect only at i .

Let l be a splittable combinatorial planar loop with n points of intersection. On splitting l at all points of intersection, we obtain a family of simple combinatorial loops $\mathcal{S}_l = \{\mathfrak{s}_1, \dots, \mathfrak{s}_{n+1}\}$ in l , which has the structure of a tree, in which \mathfrak{s}_j and $\mathfrak{s}_{j'}$ are adjacent if they share a point of intersection of l . We choose the sequence $(\mathfrak{s}_1, \dots, \mathfrak{s}_{n+1})$ to be an *adapted labelling* of \mathcal{S}_l , meaning that \mathfrak{s}_{k+1} is adjacent to at least one of $\mathfrak{s}_1, \dots, \mathfrak{s}_k$ for all $k \in \{1, \dots, n\}$. Given $T \in (0, \infty)$, a distinguished face $k \in \mathcal{F}_l$ and an adapted labelling $(\mathfrak{s}_1, \dots, \mathfrak{s}_{n+1})$ of \mathcal{S}_l , let us say that a sequence $(\gamma_1, \dots, \gamma_{n+1})$ of disjoint simple loops in \mathbb{C} around $[-\beta, \beta]$ is *admissible* if

- (a) γ_{j+1} lies in the infinite component of $\mathbb{C} \setminus \gamma_j^*$ for all $j \leq n$,
- (b) γ_j has the same orientation in \mathbb{C} as \mathfrak{s}_j has around k for all j .

For any self-intersection point i of l , we label the loops among l_i, \hat{l}_i using the left and right outgoing edges at i by $l_{i,l}$ and $l_{i,r}$ respectively. The loops $l_{i,l}$ and $l_{i,r}$ are also splittable, and the pair $\{\mathcal{S}_{l_{i,l}}, \mathcal{S}_{l_{i,r}}\}$ is a partition of \mathcal{S}_l . Write $j(i, l)$ and $j(i, r)$ for the loop labels in \mathcal{S}_l such that $\mathfrak{s}_{j(i, l)}$ and $\mathfrak{s}_{j(i, r)}$ use the left and right outgoing edges at i respectively. Let n_l be the winding number function of l , where the additive constant is chosen so that $n_l(k) = 0$. Set $\varepsilon_j = -1$ or $\varepsilon_j = 1$ according as \mathfrak{s}_j winds positively or negatively around k . Set

$$\mathcal{O}_l = \{(z_1, \dots, z_{n+1}) \in \mathbb{C}^{n+1} : z_j \neq z_{j'} \text{ for all } j, j' \text{ distinct}\}$$

and, for $a \in \Delta_l(T)$ and $z \in \mathcal{O}_l$, define

$$Q_{l,k}(a, z) = \frac{\prod_{j=1}^{n+1} \exp\{\langle n_l, a \rangle z_j + \varepsilon_j G_T(z_j)\}}{\prod_{i \in \mathcal{I}} (z_{j(i, r)} - z_{j(i, l)})},$$

where \mathcal{I} denotes, as in section 4.1, the set of points of intersection of l . Recall from Subsection 4.5 that, for all combinatorial planar loops l , there is a uniformly continuous map

$$\phi_l : \Delta_l(T) \rightarrow \mathbb{R}$$

⁷ Note that, contrary to what it is claimed in [49], $n_{i,j}$ can be even or odd.

⁸ In [35] these loops are called planar. A loop is splittable if and only if, with notations from [49], $n_{i,j} = 1$ for all window's labels i, j .

such that $\Phi_T(l) = \phi_l(a)$ for all $a \in \Delta_l(T)$ and all $l \in \mathcal{L}(a)$.

Proposition 6.6 *For all $T \in (0, \infty)$, all splittable combinatorial planar loops \mathfrak{l} with n self-intersections and equipped with a distinguished face k , all adapted labellings $(\mathfrak{s}_1, \dots, \mathfrak{s}_{n+1})$ of \mathcal{S}_l , and all admissible sequences of closed loops $(\gamma_1, \dots, \gamma_{n+1})$, we have, for all $a \in \Delta_l(T)$,*

$$\phi_l(a) = \left(\frac{1}{2\pi i} \right)^{n+1} \int_{\gamma_1} dz_1 \dots \int_{\gamma_{n+1}} dz_{n+1} \frac{\prod_{j=1}^{n+1} \exp\{\langle n_l, a \rangle z_j + \varepsilon_j G_T(z_j)\}}{\prod_{i \in \mathcal{I}} (z_{j(i,r)} - z_{j(i,l)})}. \quad (53)$$

Before proving this proposition, let us give an example with maximally winding loops, as defined in section 6.4. Any maximally winding loop \mathfrak{l}_n , with $n \geq 1$ is splittable. According to proposition 6.6, using the labelling of section 6 as in Figure 6 and choosing 0 as the distinguished face, for any $(a_0, \dots, a_n) \in \Delta_{\mathfrak{l}_n}(T)$,

$$\phi_{\mathfrak{l}_n}(a_0, a_1, \dots, a_n) = \left(\frac{1}{2\pi i} \right)^n \int_{\gamma_1} dz_1 \dots \int_{\gamma_n} dz_n \frac{\prod_{j=1}^n \exp\{(\sum_{i=j}^n a_i) z_j - G_T(z_j)\}}{(z_1 - z_2)(z_2 - z_3) \dots (z_{n-1} - z_n)} \quad (54)$$

where $(\gamma_1, \dots, \gamma_n)$ are nested clockwise-oriented simple loops in \mathbb{C} around $[-\beta, \beta]$ and g_{j+1} lies in the infinite component of $\mathbb{C} \setminus \gamma_j^*$ for all $j \leq n-1$. Recall that if l is a simple loop which divides \mathbb{S}_T into components of areas a and b , then $\Phi_T(l^n) = \phi_{\mathfrak{l}_n}(a, 0, \dots, 0, b)$. Therefore, for such a simple loop l , the last formula yields the identity

$$\Phi_l(l^n) = \left(\frac{1}{2\pi i} \right)^n \int_{\gamma_1} dz_1 \dots \int_{\gamma_n} dz_n \frac{\prod_{j=1}^n \exp\{bz_j - G_T(z_j)\}}{(z_1 - z_2)(z_2 - z_3) \dots (z_{n-1} - z_n)}. \quad (55)$$

Thanks to the following lemma and proposition 3.7, this new formula agree the first formula we obtained (11).

Lemma 6.7 *For $n \geq 2$ and $F : (\mathbb{C} \setminus [-\beta, \beta])^n \rightarrow \mathbb{C}$ a holomorphic symmetric function,*

$$\left(\frac{1}{2\pi i} \right)^n \int_{\gamma_1} dz_1 \dots \int_{\gamma_n} dz_n \frac{F(z_1, \dots, z_n)}{(z_1 - z_2)(z_2 - z_3) \dots (z_{n-1} - z_n)} = \frac{1}{2i\pi n} \int_{\gamma_1} F(z, \dots, z) dz$$

where $(\gamma_1, \dots, \gamma_n)$ are nested contours as in (54).

Proof. For $n = 2$, the formula follows from the residue theorem. The result can then be proved by induction on n , using the identity

$$\sum_{m=0}^{n-1} \frac{1}{\prod_{k=1}^{n-1} (z_{\sigma^m(k)} - z_{\sigma^m(k+1)})} = 0$$

where σ is the full cycle $(1\ 2 \dots n)$, changing variables, using the residue theorem and lastly the induction hypothesis. The details are left to the reader. \square

To prove Proposition 6.6, we will need the following technical lemma. Set

$$\Delta_{l, \mathbb{C}}(T) = \left\{ (a_k : k \in \mathcal{F}_l) : a_k \in \mathbb{C}, \sum_k a_k = T \right\}.$$

Lemma 6.8 *The map ϕ_l has an analytic extension $\Delta_{l, \mathbb{C}}(T) \rightarrow \mathbb{C}$.*

Proof. The following formula is the case $t = 1$ of (44):

$$\phi_T(n, a_0, a_*) = \phi_l(a) + \sum_{i \in \mathcal{I}} \alpha_i \int_0^1 \phi_{l_i}(a(s)) \phi_{\tilde{l}_i}(a(s)) ds \quad (56)$$

where the left-hand side is defined by (12). We see from (12) that $\phi_T(n, \cdot, \cdot)$ has an analytic extension to $\Delta_{s, \mathbb{C}}(T)$. Also, the real linear maps

$$a \mapsto \alpha : \Delta_l(T) \rightarrow \mathbb{R}^{\mathcal{I}}, \quad a \mapsto (a_0, a_*) : \Delta_l(T) \rightarrow \Delta_s(T)$$

extend to complex linear maps $\Delta_{l, \mathbb{C}}(T) \rightarrow \mathbb{C}^{\mathcal{I}}$ and $\Delta_{l, \mathbb{C}}(T) \rightarrow \Delta_{s, \mathbb{C}}(T)$. We can therefore use (56) recursively to construct the desired analytic extension of ϕ_l . \square

Proof of Proposition 6.6. Since $Q_{l,k}(a, z)$ is continuous in $z = (z_1, \dots, z_{n+1})$ on \mathcal{O}_S , analytic in a , and uniformly bounded on compacts in $\Delta_{l, \mathbb{C}}(T)$, the right-hand side of (53) is a well-defined multiple contour integral, does not depend on the order of integration, does not depend on the choice of admissible family $(\gamma_1, \dots, \gamma_{n+1})$, and defines an analytic function ψ_l on $\Delta_{l, \mathbb{C}}(T)$. Set

$$\delta_l(a) = \phi_l(a) - \psi_l(a).$$

Then δ_l is also analytic on $\Delta_{l, \mathbb{C}}(T)$ by Proposition 6.8. We will show by induction on n that $\delta_l(a) = 0$.

For $n = 0$, this follows from Proposition 2.5. Suppose inductively that the statement holds for $n - 1$ and let l be a splittable combinatorial planar loop with n intersections. Fix $i \in \mathcal{I}_l$, to be chosen later, and write k_l and k_r for the labels in $l_{i,l}$ and $l_{i,r}$ of the faces containing the face k in l . For $a \in \Delta_l(T)$, write a_l and a_r for the images of a under the natural submersions

$$\Delta_{l, \mathbb{C}}(T) \rightarrow \Delta_{l_{i,l}, \mathbb{C}}(T), \quad \Delta_{l, \mathbb{C}}(T) \rightarrow \Delta_{l_{i,r}, \mathbb{C}}(T).$$

For $(z_1, \dots, z_{n+1}) \in \mathbb{C}^{n+1}$, set

$$z_l = (z_j : s_j \in \mathcal{S}_{l_{i,l}}), \quad z_r = (z_j : s_j \in \mathcal{S}_{l_{i,r}}).$$

Then, for $a \in \Delta_{l, \mathbb{C}}(T)$ and $s \in \mathcal{S}$, we have

$$\Xi_i \langle n_s, a \rangle = \begin{cases} 1, & \text{if } s \text{ uses the right outgoing edge at } i, \\ -1, & \text{if } s \text{ uses the left outgoing edge at } i, \\ 0, & \text{otherwise.} \end{cases}$$

Hence

$$\Xi_i Q_{l,k}(a, z) = Q_{l_{i,l}, k_l}(a_l, z_l) Q_{l_{i,r}, k_r}(a_r, z_r). \quad (57)$$

Since $n \geq 1$, the tree \mathcal{S}_l has at least two leaves, and one of them, say s_m , is not the boundary of the distinguished face k . Since the labelling is adapted, there exists $p \leq m - 1$ such that s_p is adjacent to s_m . Denote by k_c the component of its complement which does not include k_∞ and set $i \in \mathcal{I}_l$ to be the intersection point s_m and s_p are sharing. The sequence $(s_1, \dots, s_{m-1}, s_{m+1}, \dots, s_n)$ is an adapted labelling of $l_{i,l}$ and the family of loops $(\gamma_1, \dots, \gamma_{m-1}, \gamma_{m+1}, \dots, \gamma_n)$ is admissible for this sequence and for the distinguished face k_l . Also, s_m is an adapted labelling of $l_{i,r}$ with admissible loop γ_m . Since the right-hand side of (57) is uniformly bounded on any compact subset of $\Delta_{l, \mathbb{C}}(T) \times \mathcal{O}_n$, we deduce that, for all $a \in \Delta_{l, \mathbb{C}}(T)$,

$$\Xi_i \psi_l(a) = \psi_{l_{i,l}}(a_l) \psi_{l_{i,r}}(a_r).$$

On the other hand, since Φ_T satisfies the Makeenko–Migdal equations, for all $a \in \Delta_l(T)$,

$$\Xi_i \phi_l(a) = \phi_{l_{i,l}}(a_l) \phi_{l_{i,r}}(a_r)$$

and this extends to $a \in \Delta_{\mathbb{I},\mathbb{C}}(T)$ by analyticity. But $\mathfrak{l}_{i,l}$ and $\mathfrak{l}_{i,r}$ are splittable and have no more than $n - 1$ points of intersection. So we have shown that, for all $a \in \Delta_{\mathbb{I},\mathbb{C}}(T)$,

$$\Xi_i \delta_{\mathfrak{l}}(a) = 0. \quad (58)$$

We check now the boundary condition of this equation. Since \mathfrak{l} is splittable, there is a splittable loop $\tilde{\mathfrak{l}}$, with exactly $n - 1$ intersections, an affine map

$$\iota_c : \Delta_{\mathfrak{l}}(T) \cap \{a : a_{k_c} = 0\} \rightarrow \Delta_{\tilde{\mathfrak{l}}}(T)$$

and a distinguished face $\tilde{k} \in \mathcal{F}_{\tilde{\mathfrak{l}}}$ such that, for any $a \in \Delta_{\mathfrak{l}}(T)$ with $a_{k_c} = 0$,

$$\mathfrak{l}(a) \cap \tilde{\mathfrak{l}}(\iota_c(a)) \neq \emptyset$$

and $\iota_c(a)_{\tilde{k}} = 0$ if and only if $a_k = 0$. Moreover, for all $a \in \Delta_{\mathfrak{l}}(T)$,

$$\phi_{\mathfrak{l}}(a) = \phi_{\tilde{\mathfrak{l}}}(\iota_c(a)). \quad (59)$$

Furthermore, by analyticity of $\phi_{\mathfrak{l}}$ and $\phi_{\tilde{\mathfrak{l}}}$, this equality holds true for all $a \in \Delta_{\mathbb{I},\mathbb{C}}(T)$ with $a_{k_c} = 0$. Let $\nu \in \mathbb{Z}^{\mathcal{F}_{\mathfrak{l}}}$ be the vector with $\nu_{k_c} = 1$ which is proportional to Ξ_i , viewed as an element of $\{1_{\mathcal{F}_{\mathfrak{l}}}\}^{\perp} \cap \mathbb{R}^{\mathcal{F}_{\mathfrak{l}}}$, where i is the only vertex adjacent to F_c . Then, by (58), for all $a \in \Delta_{\mathbb{I},\mathbb{C}}(T)$,

$$\psi_{\mathfrak{l}}(a) = \psi_{\mathfrak{l}}(a - a_{k_c} \nu).$$

As $a - a_{k_c} \nu \in \Delta_{\mathbb{I},\mathbb{C}}(T) \cap \{a : a_{k_c} = 0\}$, by (59), in order to conclude, it is sufficient to show that, for all $a \in \Delta_{\mathbb{I},\mathbb{C}}(T)$ with $a_{k_c} = 0$,

$$\phi_{\mathfrak{l}}(a) = \phi_{\tilde{\mathfrak{l}}}(\iota_c(a)).$$

For such a vector a and for $z \in \mathcal{O}_n$, set $\tilde{z} = (z_j : j \neq m)$. Then

$$Q_{\mathfrak{l},k}(a, z) = Q_{\tilde{\mathfrak{l}},\tilde{k}}(a, \tilde{z}) \frac{\varepsilon_m e^{\varepsilon_m G_T(z_m)}}{z_m - z_p}.$$

For $a \in \Delta_{\mathbb{I},\mathbb{C}}(T)$, the only singularity of $z_m \in \mathbb{C} \setminus [-\beta, \beta] \mapsto Q_{\mathfrak{l},k}(a, z)$ is at z_p . Since the family of loops $(\gamma_1, \dots, \gamma_{n+1})$ is admissible, by deforming γ_m , we can assume that the bounded component of $\mathbb{C} \setminus \gamma_m$ contains all γ_j with $j \neq m$. Then, for all $\tilde{z} \in \mathcal{O}_{n-1}$,

$$\frac{1}{2\pi i} \int_{\gamma_m} Q_{\mathfrak{l},k}(a, z) dz_{s_o} = Q_{\tilde{\mathfrak{l}},\tilde{k}}(a, \tilde{z}) \frac{1}{2\pi i} \int_C \frac{e^{\varepsilon_m G_T(z_m)}}{z_m - z_p} dz_m$$

with C an anticlockwise circle with centre 0, whose interior contains all contours $(\gamma_j : j \neq m)$. Since $G_T(z) \sim 1/z$ as $z \rightarrow \infty$, it follows that

$$\frac{1}{2\pi i} \int_C \frac{e^{\varepsilon_m G_T(z_m)}}{z_m - z_p} dz_m = -\frac{1}{2\pi i} \int_{1/C} \frac{e^{\varepsilon_m G_T(1/y)}}{y(1 - yz_p)} dy = 1.$$

Therefore, performing the integration in $\phi_{\mathfrak{l}}(a)$ first with respect to z_m , we obtain, when $a \in \Delta_{\mathbb{I},\mathbb{C}}(T)$, with $a_{k_c} = 0$,

$$\begin{aligned} \phi_{\mathfrak{l}}(a) &= \left(\frac{1}{2\pi i} \right)^n \int_{z_j \in \gamma_j, \text{ for } j \neq m} Q_{\tilde{\mathfrak{l}},\tilde{k}}(a, (z_s)_{s \in \mathcal{S}_{\mathfrak{l}} \setminus \{s_m\}}) \prod_{j \neq m} dz_j \\ &= \phi_{\tilde{\mathfrak{l}}}(\iota_c(a)). \end{aligned}$$

□

Acknowledgments. The authors wish to thank Natasha Blitvić, Guillaume Cébron, Franck Gabriel, Thierry Lévy, and Mylène Maïda for several motivating and fruitful discussions about this project. They are grateful also to Franck Gabriel for his comments on a first version of this work, and to referees whose comments led to improvements in the final version.

References

1. Michael Anshelevich and Ambar N. Sengupta. Quantum free Yang-Mills on the plane. *J. Geom. Phys.*, 62(2):330–343, 2012.
2. Thomas F. Banchoff and William F. Pohl. A generalization of the isoperimetric inequality. *J. Differential Geometry*, 6:175–192, 1971/72.
3. G. Ben Arous, A. Dembo, and A. Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1):1–67, 2001.
4. Philippe Biane. Free Brownian motion, free stochastic calculus and random matrices. In *Free probability theory (Waterloo, ON, 1995)*, volume 12 of *Fields Inst. Commun.*, pages 1–19. Amer. Math. Soc., Providence, RI, 1997.
5. Philippe Biane and Roland Speicher. Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. *Probab. Theory Related Fields*, 112(3):373–409, 1998.
6. D. V. Boulatov. Wilson loop on a sphere. *Modern Phys. Lett. A*, 9(4):365–374, 1994.
7. A. Boutet de Monvel and M. V. Shcherbina. On free energy in two-dimensional $U(n)$ -gauge field theory on the sphere. *Teoret. Mat. Fiz.*, 115(3):389–401, 1998.
8. Guillaume Cébron, Antoine Dahlqvist, and Franck Gabriel. The generalized master fields. *J. Geom. Phys.*, 119:34–53, 2017.
9. Sourav Chatterjee. Rigorous solution of strongly coupled $so(n)$ lattice gauge theory in the large n limit. *Communications in Mathematical Physics*, 366(1):203–268, Feb 2019.
10. S. Chatterjee and J. Jafarov. The $1/n$ expansion for $so(n)$ lattice gauge theory at strong coupling. [arXiv 1604.04777](#), 2016.
11. Sourav Chatterjee. The leading term of the Yang–Mills free energy. *J. Funct. Anal.*, 271(10):2944–3005, 2016.
12. Benoît Collins, Antoine Dahlqvist, and Todd Kemp. The spectral edge of unitary brownian motion. *Probability Theory and Related Fields*, Feb 2017.
13. Antoine Dahlqvist. Free Energies and Fluctuations for the Unitary Brownian Motion. *Comm. Math. Phys.*, 348(2):395–444, 2016.
14. Jean-Marc Daul and Vladimir A. Kazakov. Wilson loop for large N Yang-Mills theory on a two-dimensional sphere. *Phys. Lett. B*, 335(3-4):371–376, 1994.
15. Michael R. Douglas and Vladimir A. Kazakov. Large n phase transition in continuum qcd2. *Physics Letters B*, 319(1):219 – 230, 1993.
16. B. K. Driver. A Functional Integral Approaches to the Makeenko-Migdal Equations. *ArXiv e-prints*, September 2017.
17. Bruce K. Driver. YM_2 : continuum expectations, lattice convergence, and lassos. *Comm. Math. Phys.*, 123(4):575–616, 1989.
18. Bruce K. Driver, Franck Gabriel, Brian C. Hall, and Todd Kemp. The Makeenko-Migdal equation for Yang-Mills theory on compact surfaces. *Comm. Math. Phys.*, 352(3):967–978, 2017.
19. Bruce K. Driver, Brian C. Hall, and Todd Kemp. Three proofs of the Makeenko-Migdal equation for Yang-Mills theory on the plane. *Comm. Math. Phys.*, 351(2):741–774, 2017.
20. D. Féral. On large deviations for the spectral measure of discrete Coulomb gas. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 19–49. Springer, Berlin, 2008.
21. Dana S. Fine. Quantum Yang-Mills on a Riemann surface. *Comm. Math. Phys.*, 140(2):321–338, 1991.
22. Peter J. Forrester, Satya N. Majumdar, and Grégory Schehr. Non-intersecting Brownian walkers and Yang-Mills theory on the sphere. *Nuclear Phys. B*, 844(3):500–526, 2011.
23. Peter J. Forrester, Satya N. Majumdar, and Grégory Schehr. Erratum to “Non-intersecting Brownian walkers and Yang-Mills theory on the sphere”. *Nuclear Phys. B*, 857(3):424–427, 2012.
24. F. Gabriel. A combinatorial theory of random matrices III: random walks on $\mathfrak{S}(N)$, ramified coverings and the $\mathfrak{S}(\infty)$ Yang-Mills measure. *ArXiv e-prints*, October 2015.
25. F. D. Gakhov. *Boundary value problems*. Dover Publications, Inc., New York, 1990. Translated from the Russian, Reprint of the 1966 translation.
26. Rajesh Gopakumar. The master field in generalised QCD_2 . *Nuclear Phys. B*, 471(1-2):246–260, 1996.
27. Rajesh Gopakumar and David J. Gross. Mastering the master field. *Nuclear Phys. B*, 451(1-2):379–415, 1995.
28. David J. Gross and Andrei Matytsin. Some properties of large- N two-dimensional Yang-Mills theory. *Nuclear Phys. B*, 437(3):541–584, 1995.
29. Leonard Gross, Christopher King, and Ambar Sengupta. Two-dimensional Yang-Mills theory via stochastic differential equations. *Ann. Physics*, 194(1):65–112, 1989.
30. Alice Guionnet and Mylène Maïda. Character expansion method for the first order asymptotics of a matrix integral. *Probab. Theory Related Fields*, 132(4):539–578, 2005.
31. B.C. Hall. The large- N limit for two-dimensional Yang–Mills theory. [arXiv 1705.07808](#), 2017.
32. Kurt Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.
33. Iain Johnstone and David Siegmund. On Hotelling’s formula for the volume of tubes and Naiman’s inequality. *Ann. Statist.*, 17(1):184–194, 1989.
34. Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. II*, volume 100 of *Pure and Applied Mathematics*. Academic Press, Inc., Orlando, FL, 1986. Advanced theory.

35. V. A. Kazakov. Wilson loop average for an arbitrary contour in two-dimensional $U(N)$ gauge theory. *Nuclear Phys. B*, 179(2):283–292, 1981.
36. V. A. Kazakov and I. K. Kostov. Nonlinear strings in two-dimensional $U(\infty)$ gauge theory. *Nuclear Phys. B*, 176(1):199–215, 1980.
37. Anthony W. Knap. *Lie groups beyond an introduction*, volume 140 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, second edition, 2002.
38. Derek F. Lawden. *Elliptic functions and applications*, volume 80 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1989.
39. Thierry Lévy. Yang–Mills measure on compact surfaces. *Mem. Amer. Math. Soc.*, 166(790):xiv+122, 2003.
40. Thierry Lévy. Schur–Weyl duality and the heat kernel measure on the unitary group. *Adv. Math.*, 218(2):537–575, 2008.
41. Thierry Lévy. Two-dimensional Markovian holonomy fields. *Astérisque*, 329:172, 2010.
42. Thierry Lévy. The master field on the plane. *Astérisque*, 388:IX+201, 2017.
43. Thierry Lévy and Mylène Maïda. Central limit theorem for the heat kernel measure on the unitary group. *J. Funct. Anal.*, 259(12):3163–3204, 2010.
44. Thierry Lévy and Mylène Maïda. On the Douglas–Kazakov phase transition. Weighted potential theory under constraint for probabilists. In *Modélisation Aléatoire et Statistique—Journées MAS 2014*, volume 51 of *ESAIM Proc. Surveys*, pages 89–121. EDP Sci., Les Ulis, 2015.
45. Karl Liechty and Dong Wang. Nonintersecting Brownian motions on the unit circle. *Ann. Probab.*, 44(2):1134–1211, 2016.
46. Y. M. Makeenko and Migdal A. A. Exact equation for the loop average in multicolor QCD. *Phys. Lett. B*, 88:135–137, 1979.
47. Daniel Q. Naiman. Conservative confidence bands in curvilinear regression. *Ann. Statist.*, 14(3):896–906, 1986.
48. Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.
49. B. Rusakov. Wilson loops in large N QCD on a sphere. *Phys. Lett. B*, 329:338–344, 1994.
50. Ambar Sengupta. Gauge theory on compact surfaces. *Mem. Amer. Math. Soc.*, 126(600):viii+85, 1997.
51. Ambar N. Sengupta. The large- N Yang–Mills field on the plane and free noise. In *Geometric methods in physics*, volume 1079 of *AIP Conf. Proc.*, pages 121–132. Amer. Inst. Phys., Melville, NY, 2008.
52. I. M. Singer. On the master field in two dimensions. In *Functional analysis on the eve of the 21st century, Vol. 1 (New Brunswick, NJ, 1993)*, volume 131 of *Progr. Math.*, pages 263–281. Birkhäuser Boston, Boston, MA, 1995.
53. M. Takesaki. *Theory of operator algebras. I*, volume 124 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2002. Reprint of the first (1979) edition, Operator Algebras and Non-commutative Geometry, 5.
54. Edward Witten. On quantum gauge theories in two dimensions. *Comm. Math. Phys.*, 141(1):153–209, 1991.
55. Edward Witten. Two-dimensional gauge theories revisited. *J. Geom. Phys.*, 9(4):303–368, 1992.
56. Feng Xu. A random matrix model from two-dimensional Yang–Mills theory. *Comm. Math. Phys.*, 190(2):287–307, 1997.

Communicated by name